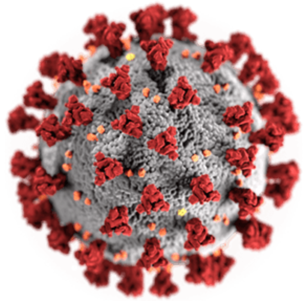
EECS 498-007 / 598-005 Deep Learning for Computer Vision

Lecture 1: Introduction

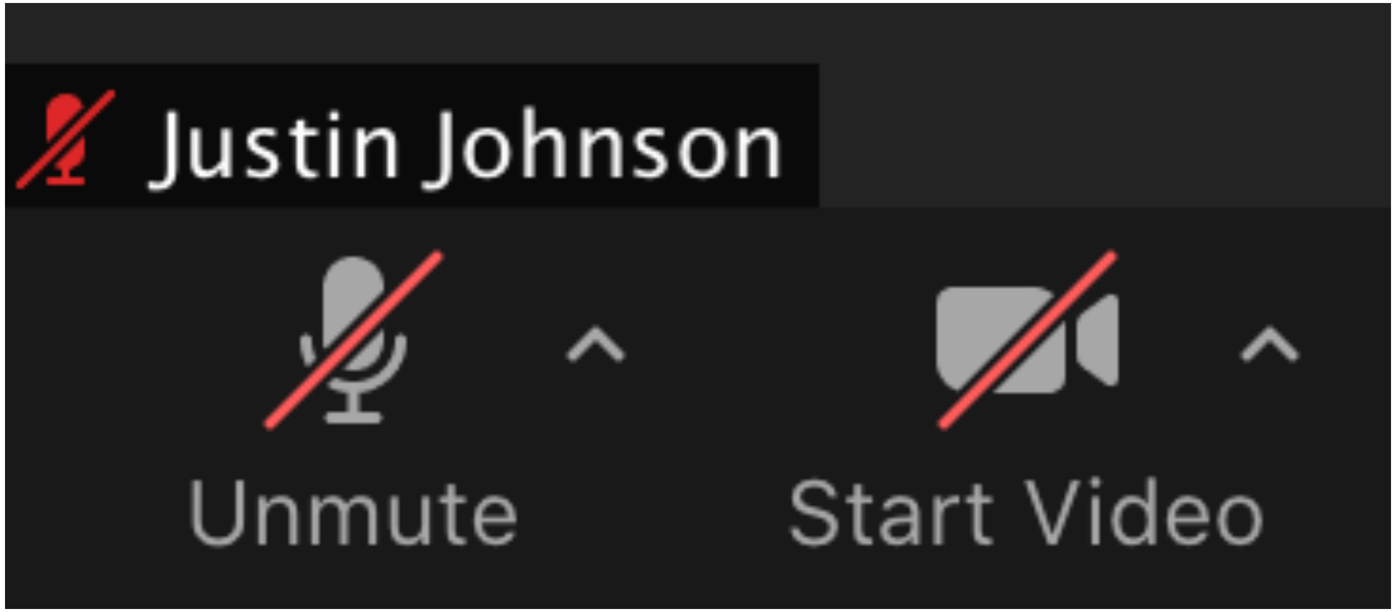
COVID-19 Edition



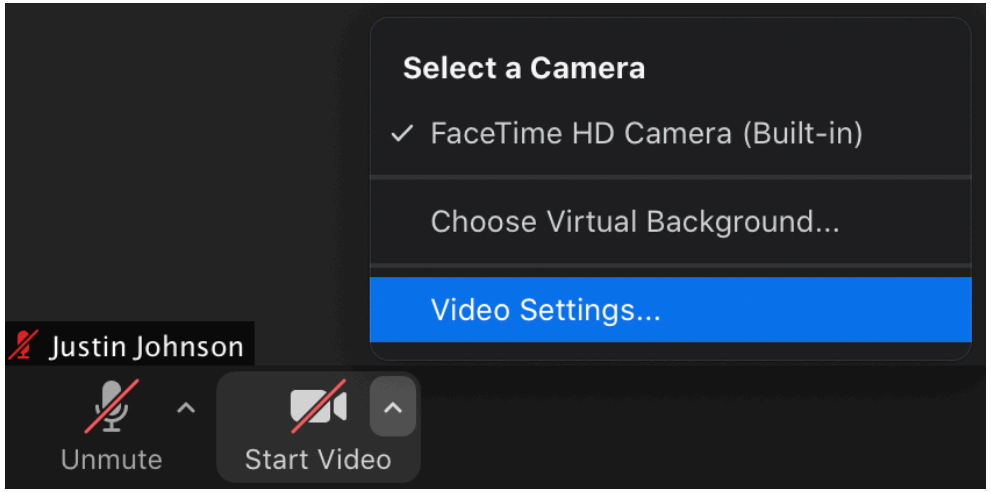
EECS 498-007 / 598-005
Deep Learning for Computer Vision

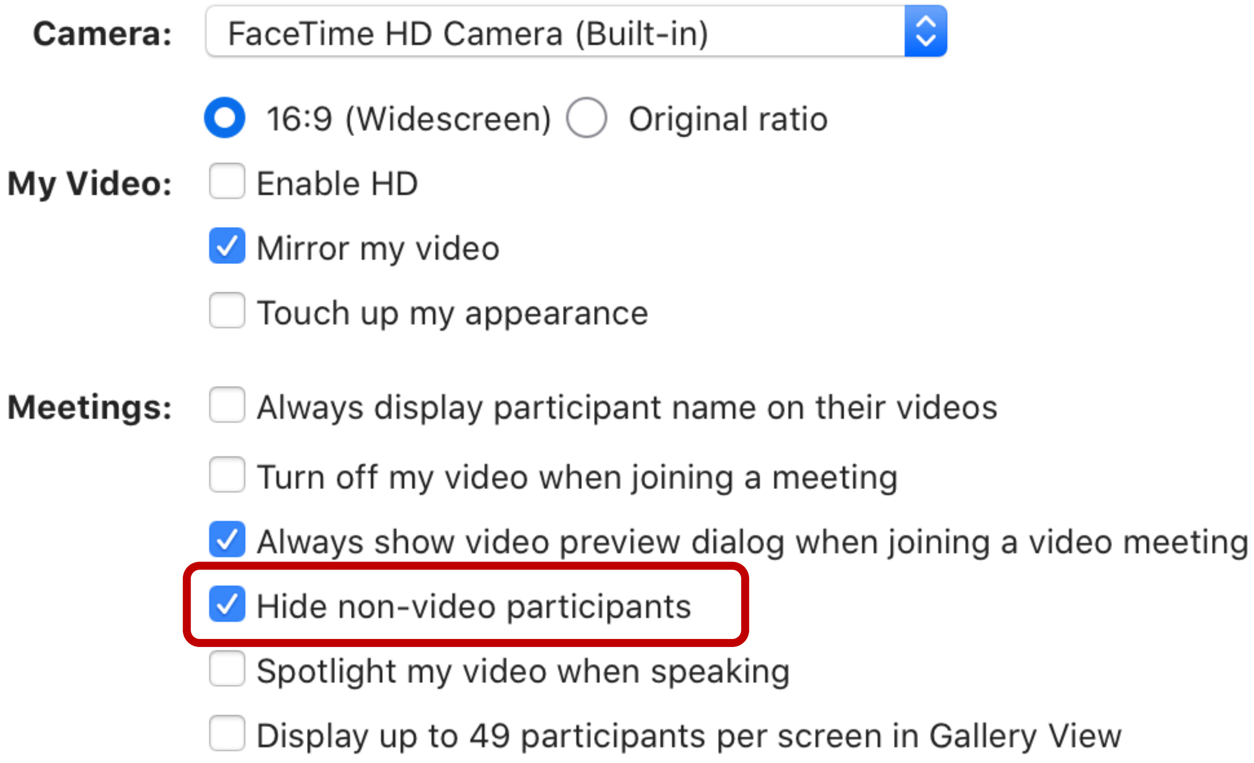
Lecture 1: Introduction

Zoom Logistics: Muted, Camera Off



Zoom Logistics: Hide non-video participants




Camera: FaceTime HD Camera (Built-in)

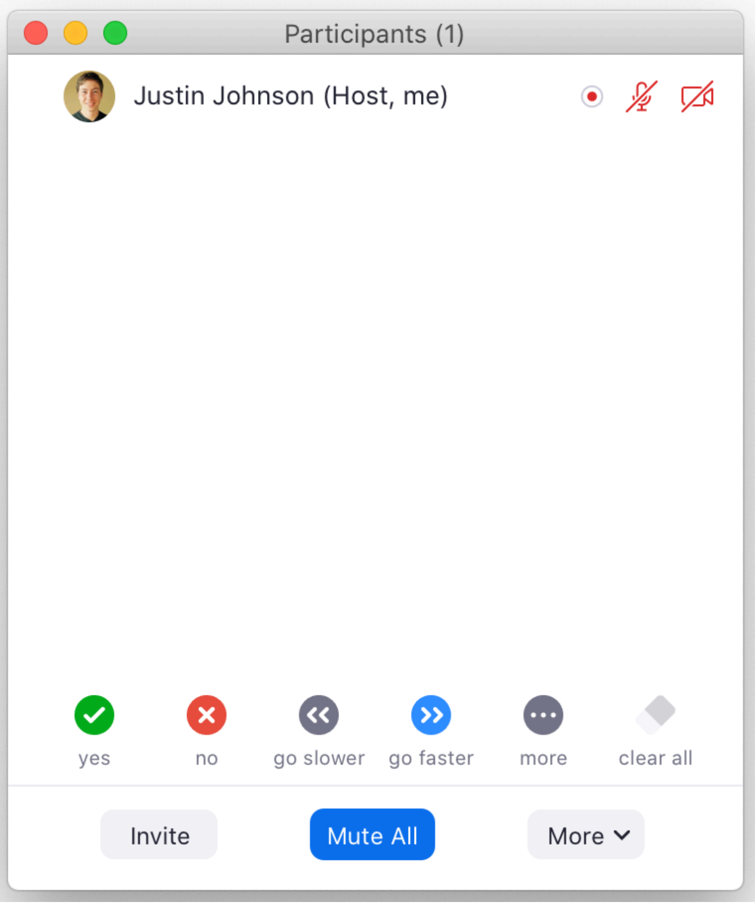
● 16:9 (Widescreen) ○ Original ratio

My Video:
- □ Enable HD
- ☑ Mirror my video
- □ Touch up my appearance

Meetings:
- □ Always display participant name on their videos
- □ Turn off my video when joining a meeting
- ☑ Always show video preview dialog when joining a video meeting
- ☑ Hide non-video participants
- □ Spotlight my video when speaking
- □ Display up to 49 participants per screen in Gallery View

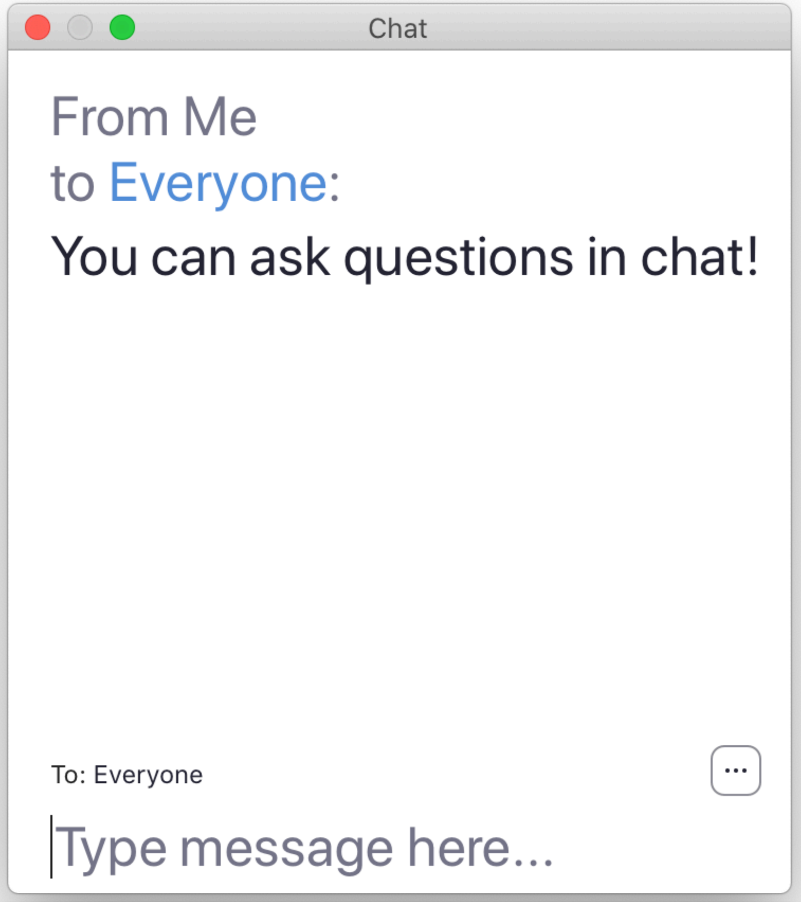
Zoom Logistics: Quick Poll

- You can use responses in ”participants” panel for quick polls



Zoom Logistics: Asking Questions

Option 1: Ask questions in chat

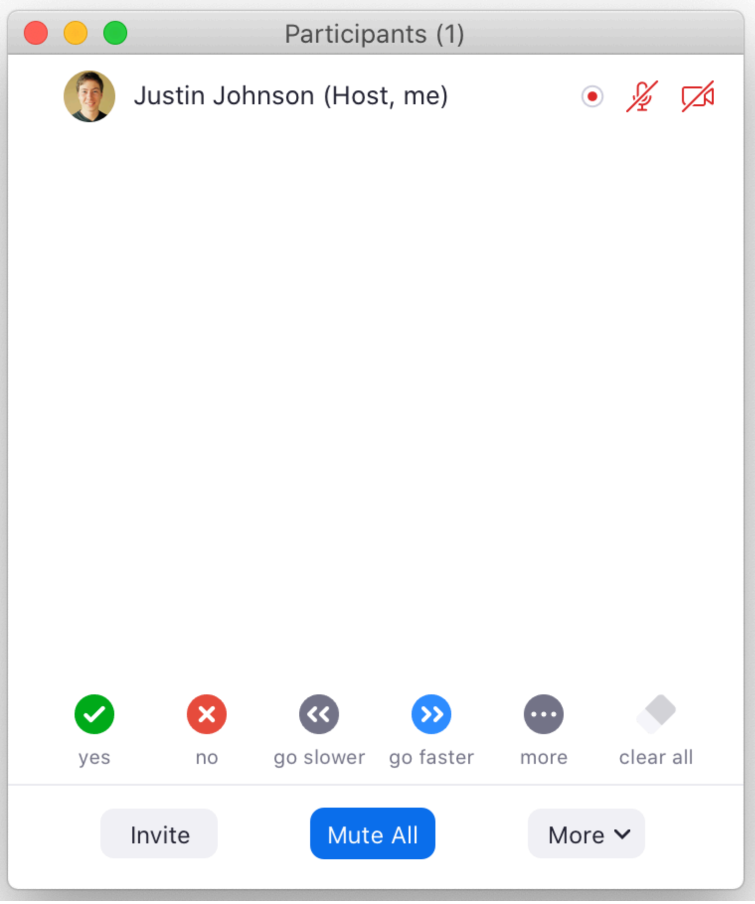


From Me
to Everyone:
You can ask questions in chat!

To: Everyone

Type message here...

Option 2: "Raise Hand" in participants

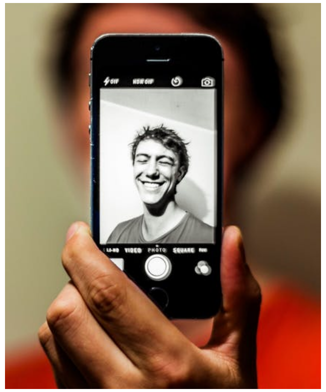


Deep Learning for Computer Vision

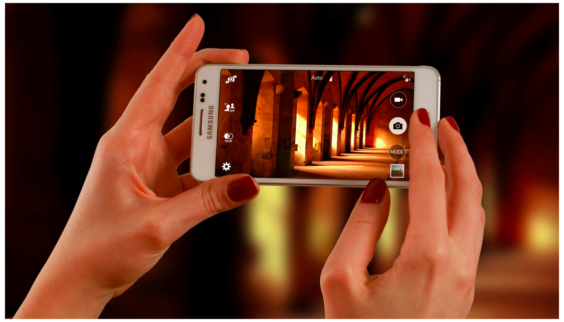
Deep Learning for Computer Vision

Building artificial systems that process, perceive, and reason about visual data

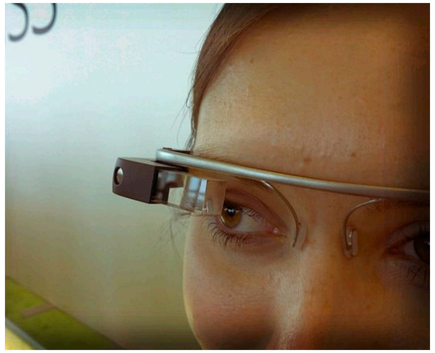
Computer Vision is everywhere!



Left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain



Left to right:
Image is free to use
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0
Image is CC0 1.0 public domain



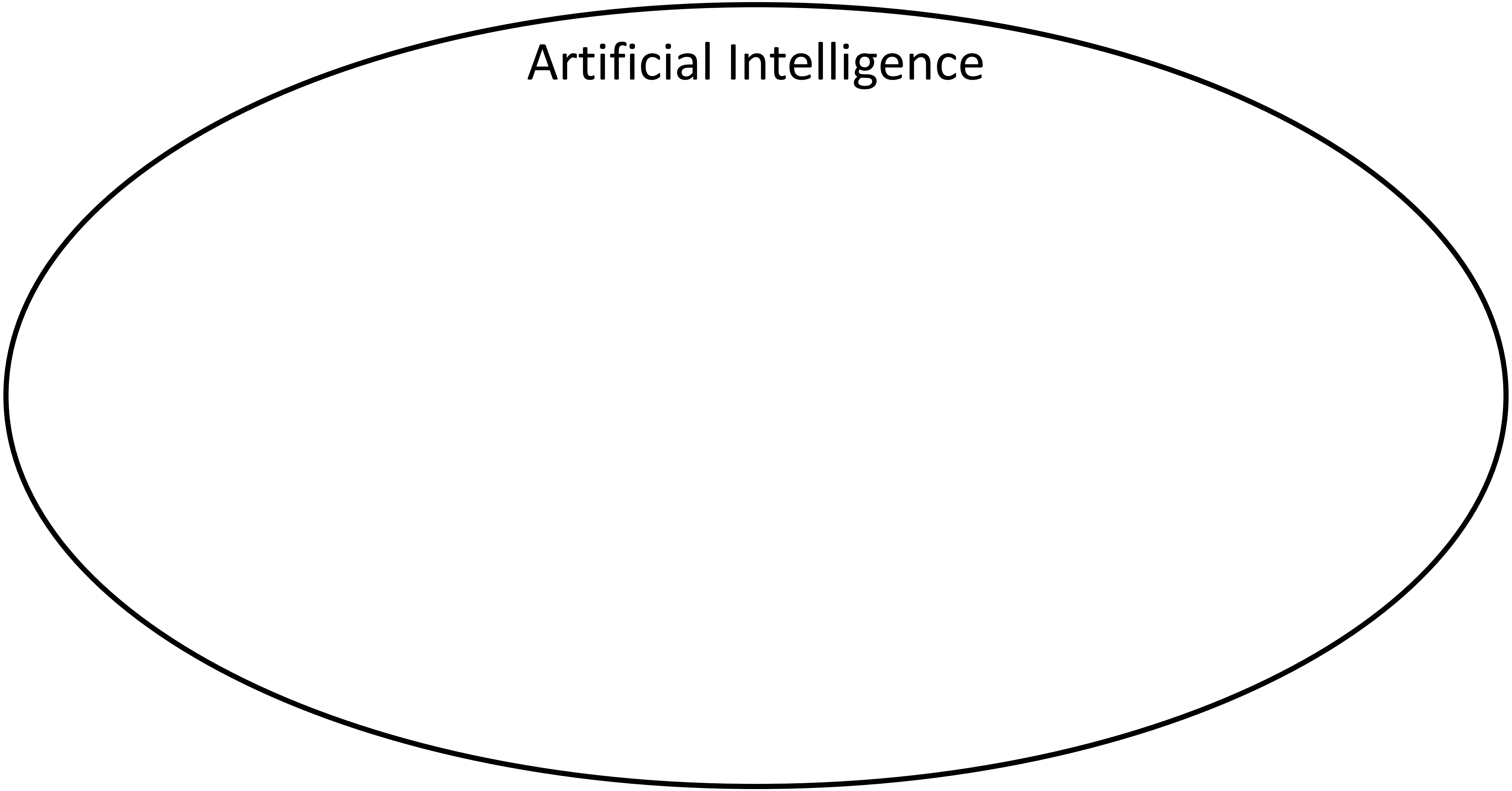
Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image is licensed under CC-BY 2.0; changes made

Deep Learning for Computer Vision

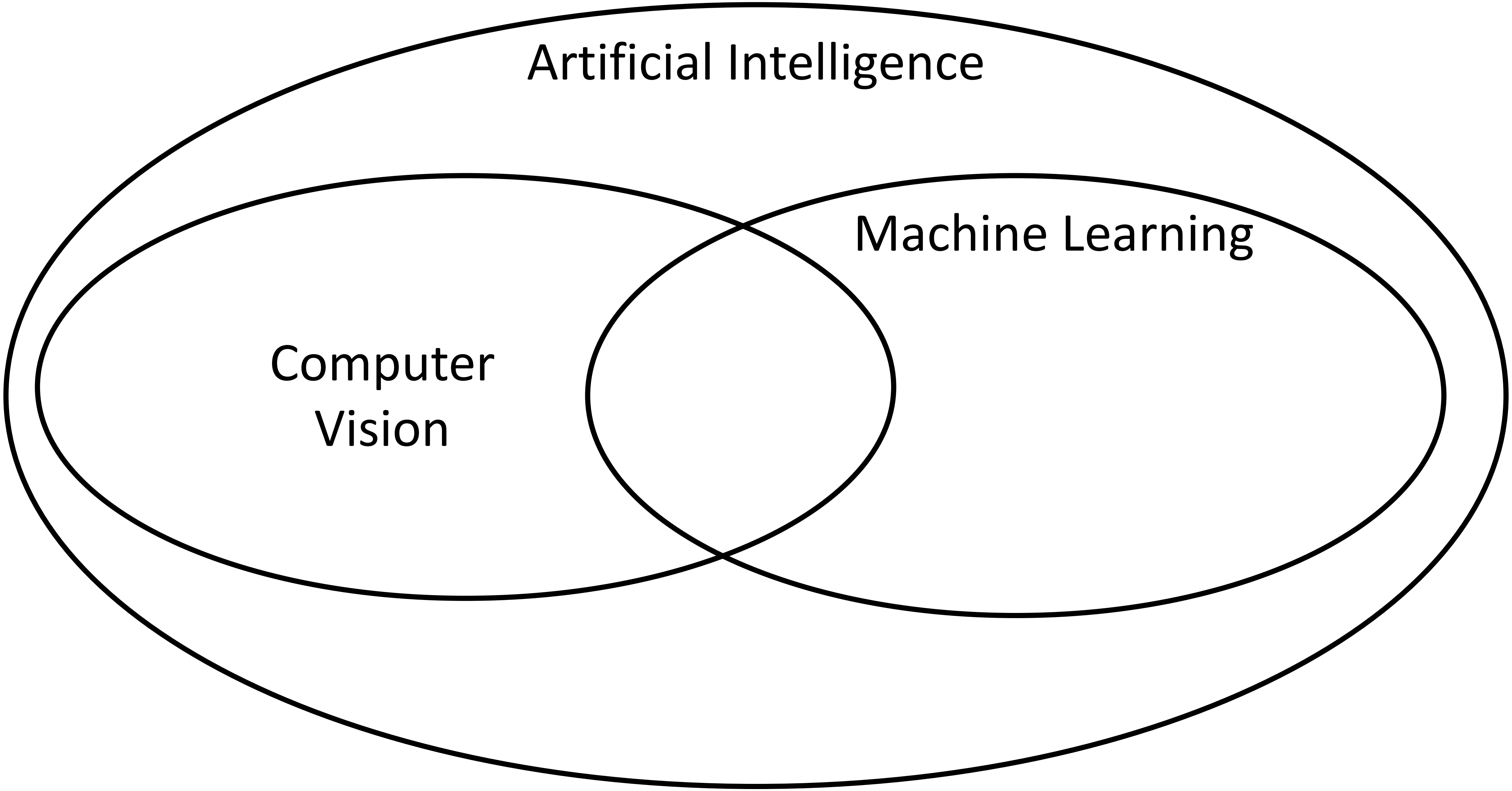
Building artificial systems that learn from data and experience

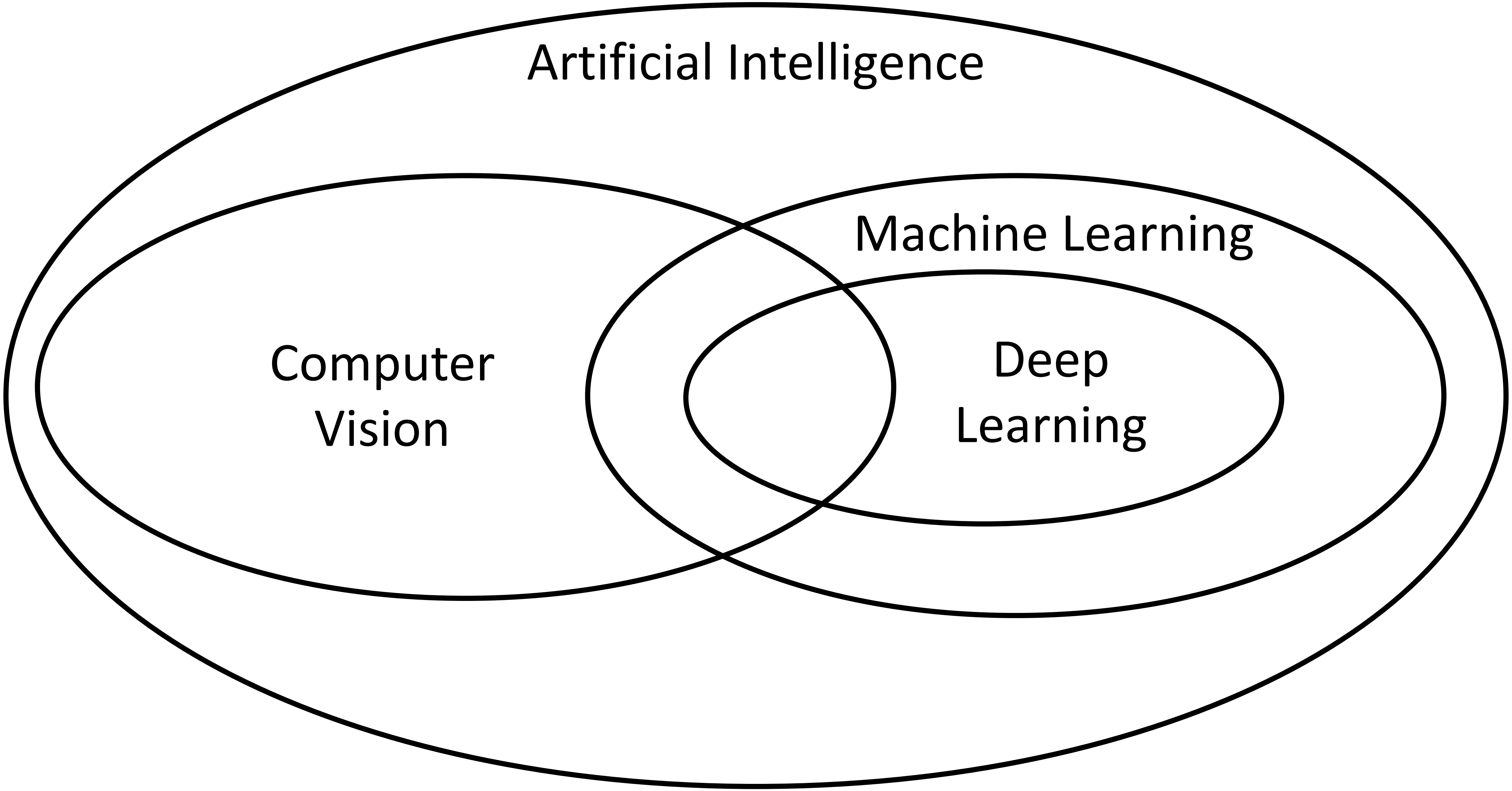
Deep Learning for Computer Vision

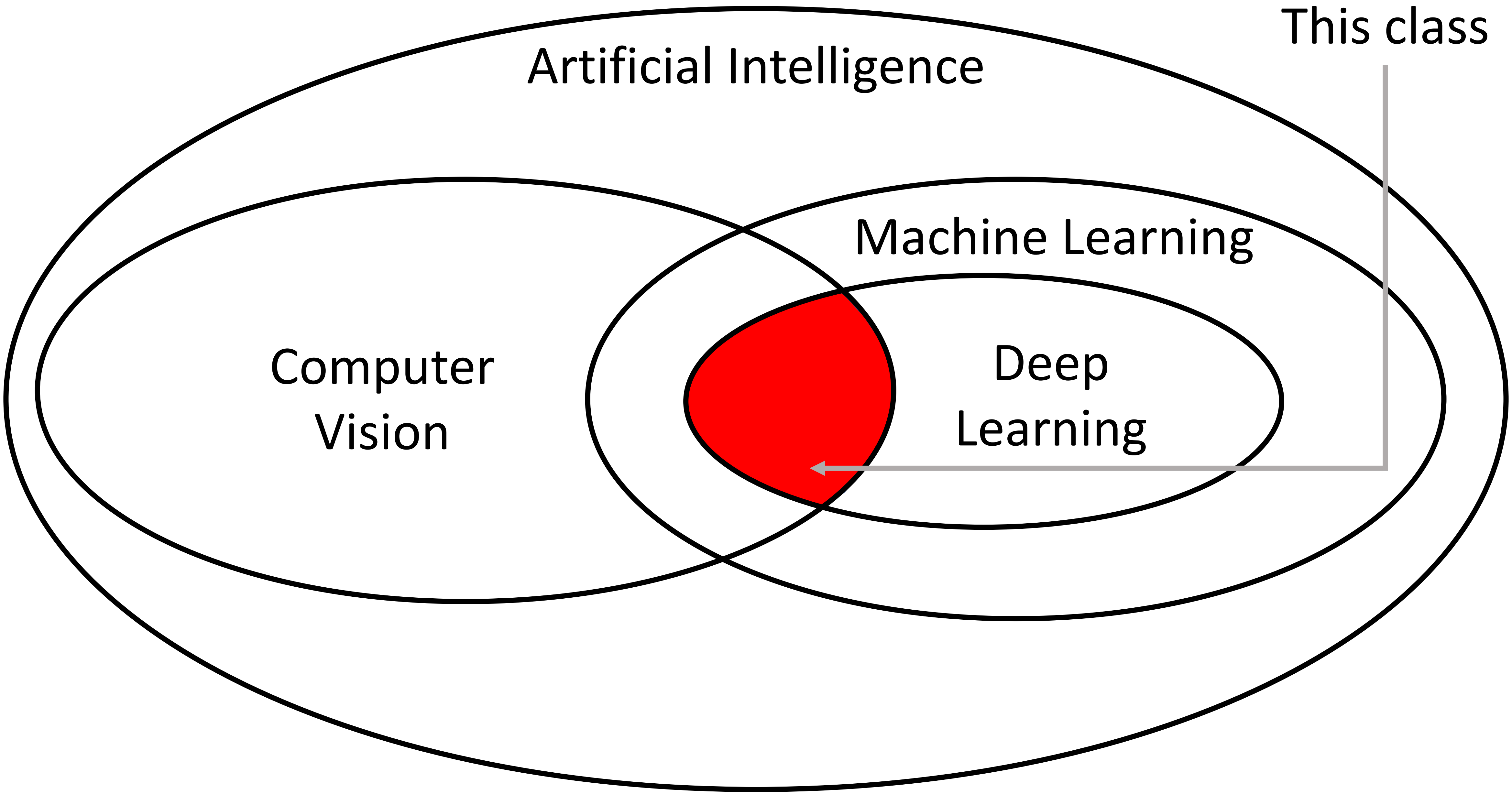
Hierarchical learning algorithms with many “layers”, (very) loosely inspired by the brain

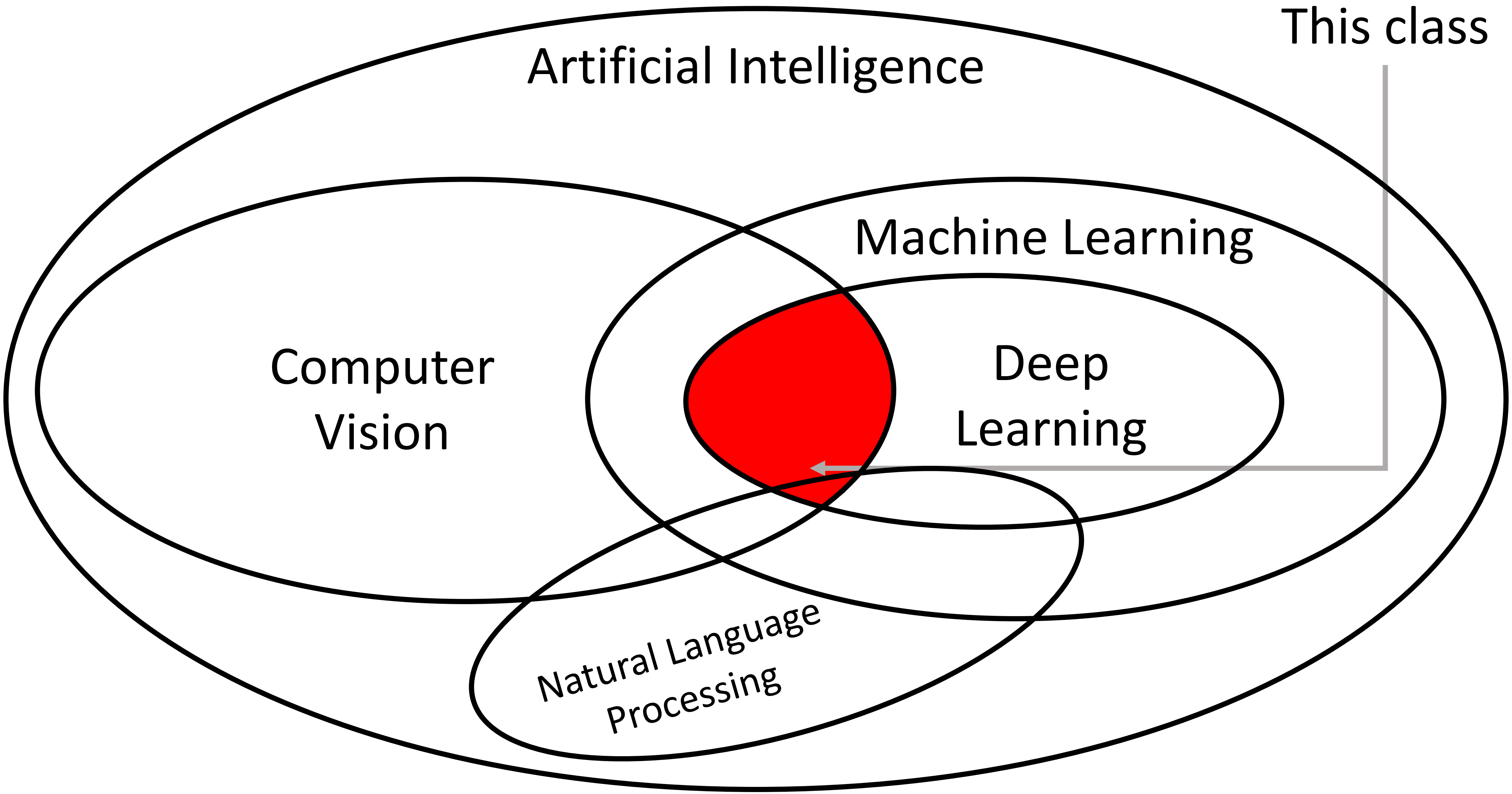


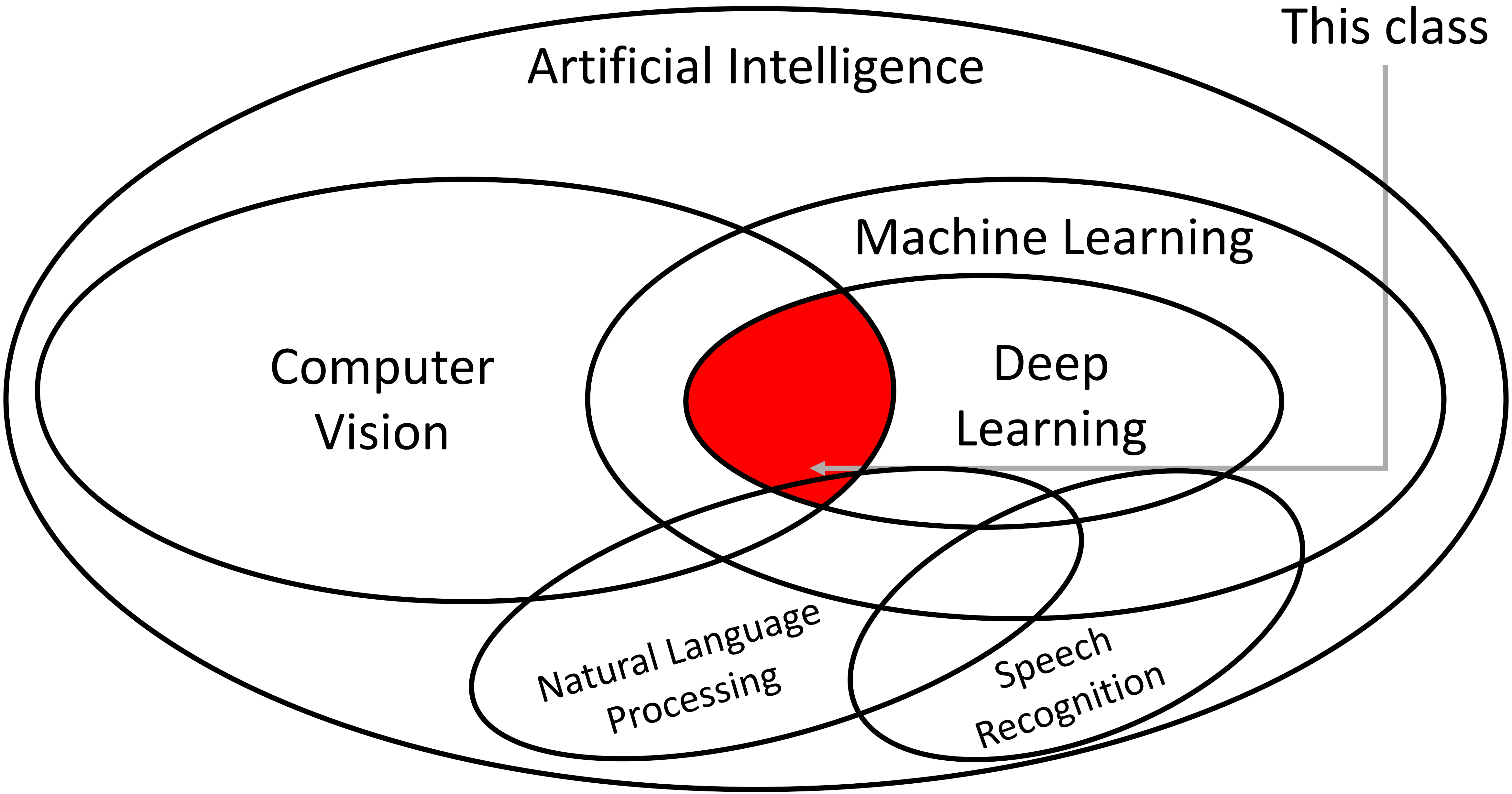
Artificial Intelligence

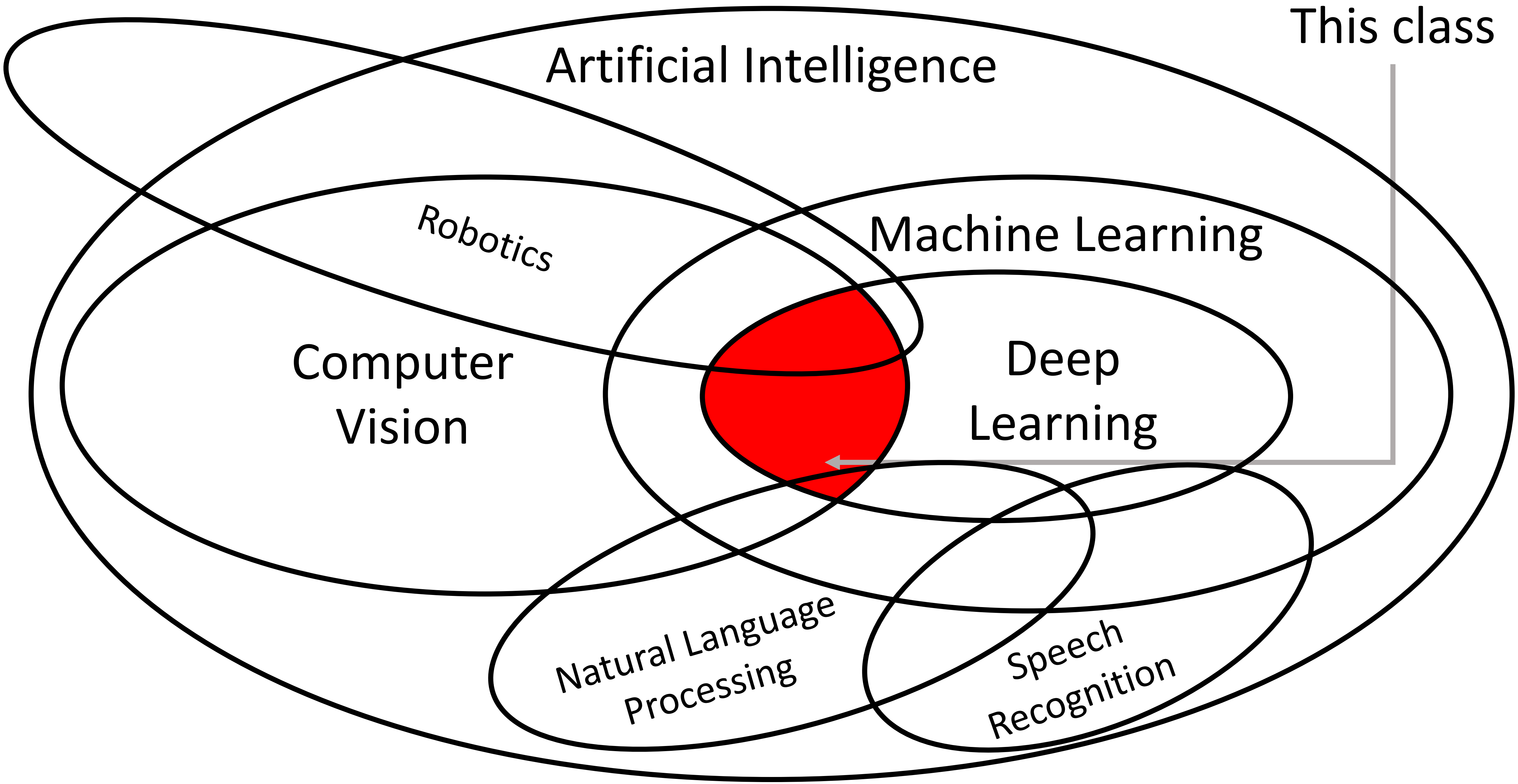

Artificial Intelligence
Machine Learning
Computer Vision


Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning


This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning


This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing


This class
Artificial Intelligence
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition


This class
Artificial Intelligence
Robotics
Machine Learning
Computer Vision
Deep Learning
Natural Language Processing
Speech Recognition

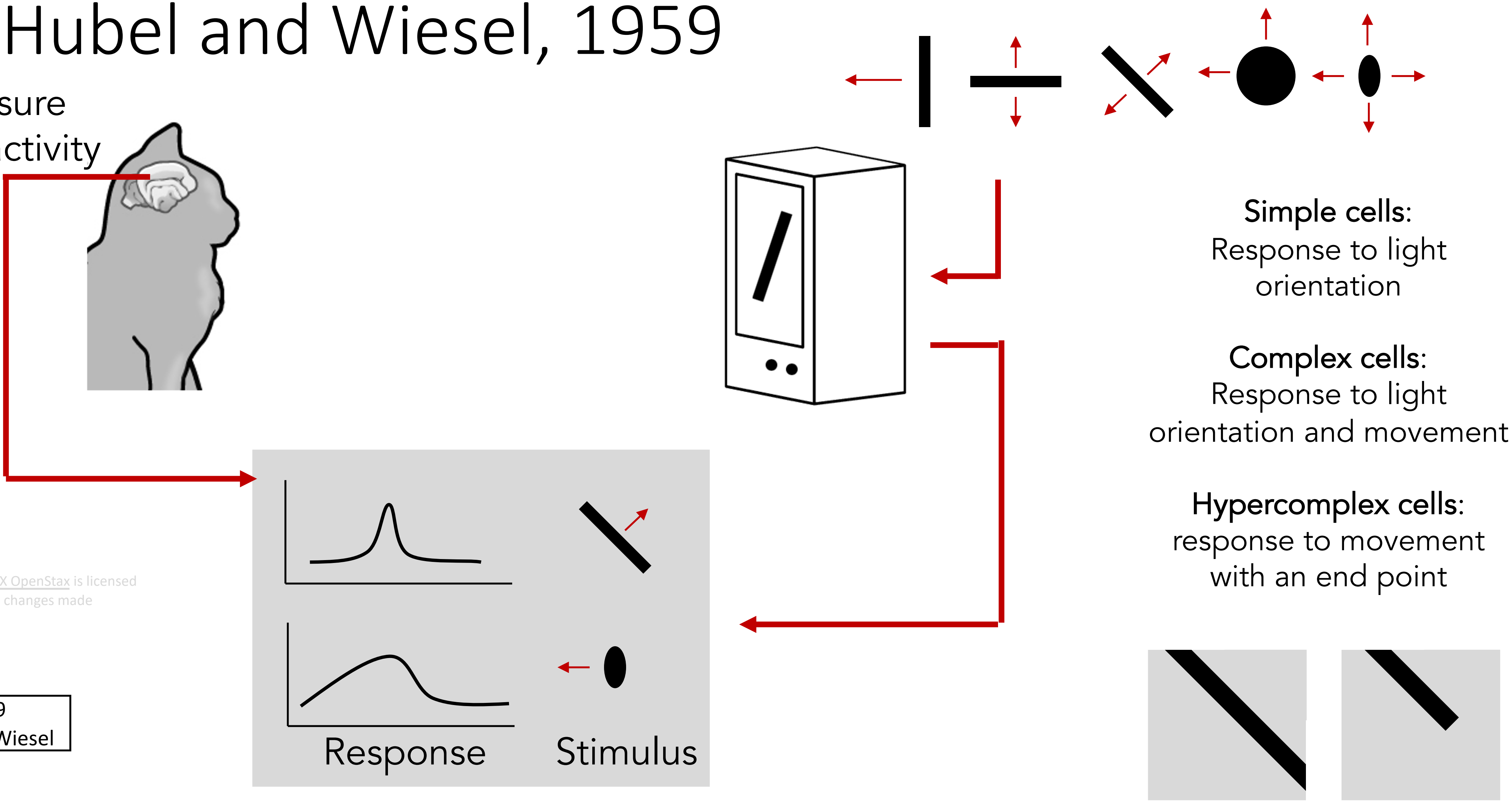
Today’s Agenda

- A brief history of computer vision and deep learning

- Course overview and logistics

Hubel and Wiesel, 1959

Measure brain activity



Simple cells: Response to light orientation

Complex cells: Response to light orientation and movement

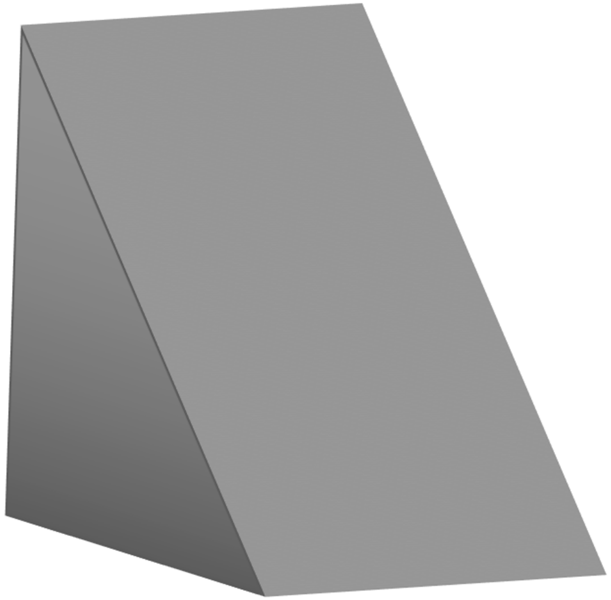
Hypercomplex cells: response to movement with an end point

No response

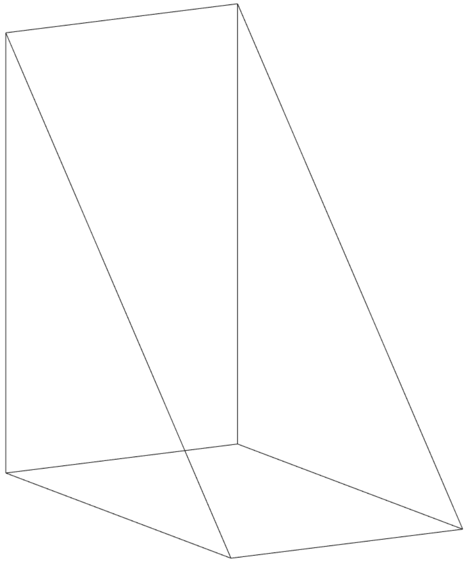
Cat image by CNX OpenStax is licensed under CC BY 4.0; changes made

1959
Hubel & Wiesel

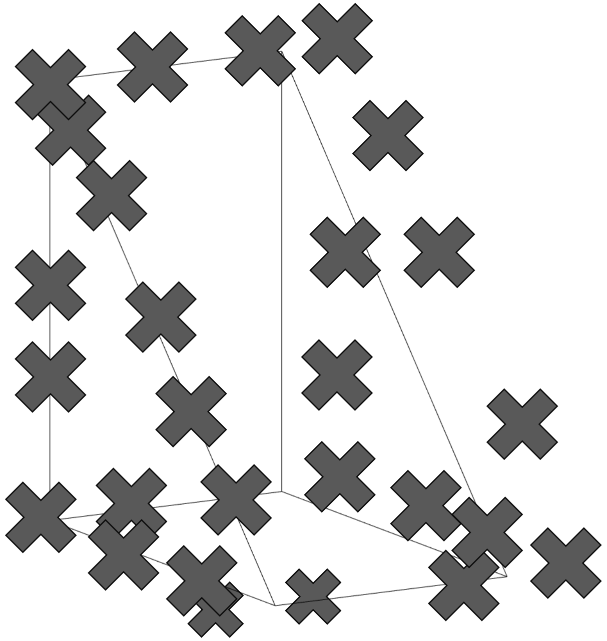
Larry Roberts, 1963



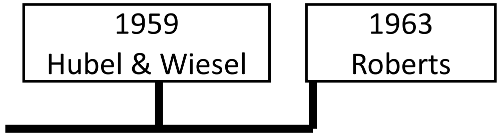
(a) Original picture



(b) Differentiated picture

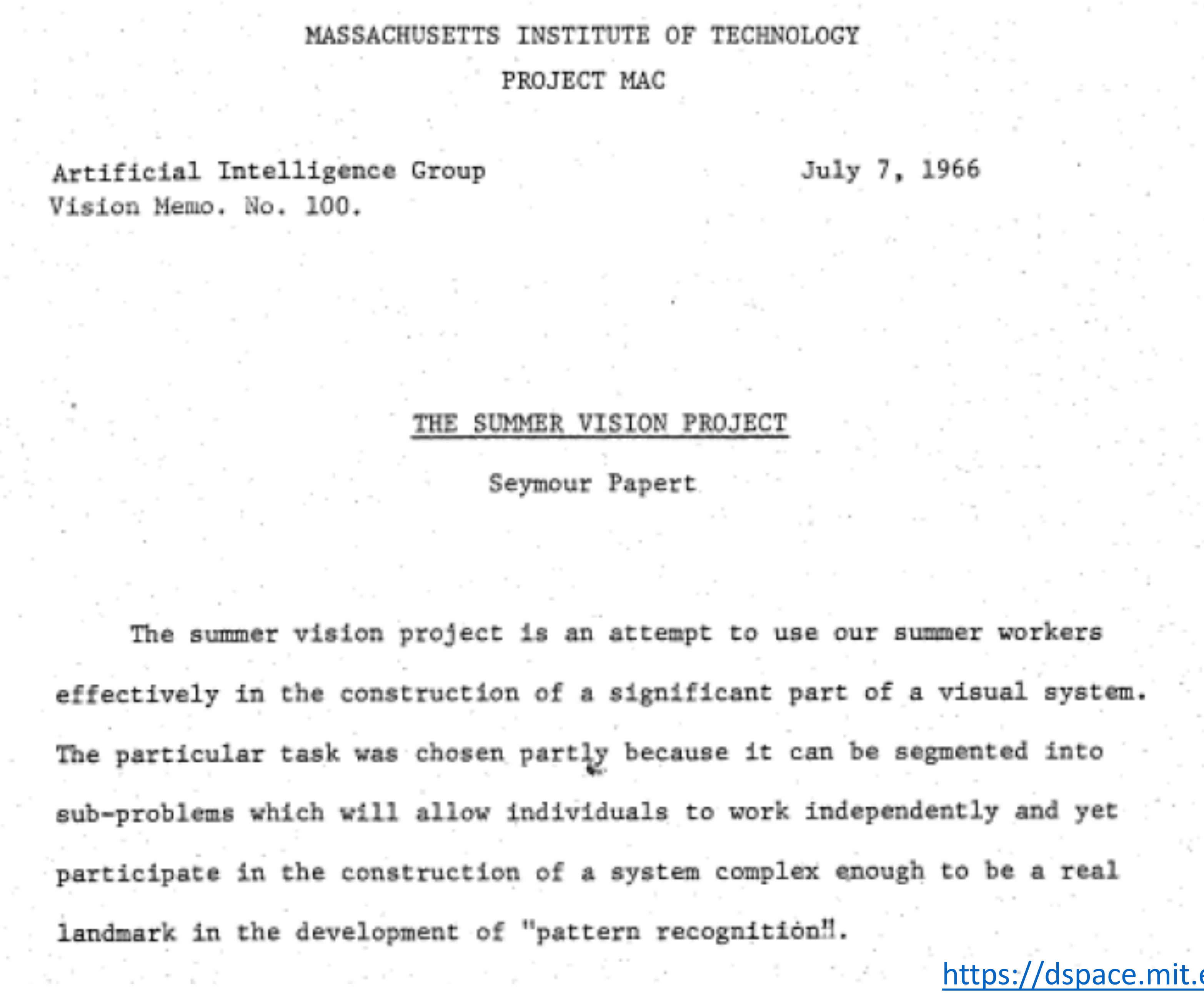


(c) Feature points selected



1959
Hubel & Wiesel

1963
Roberts



MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

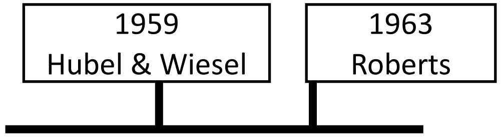
Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

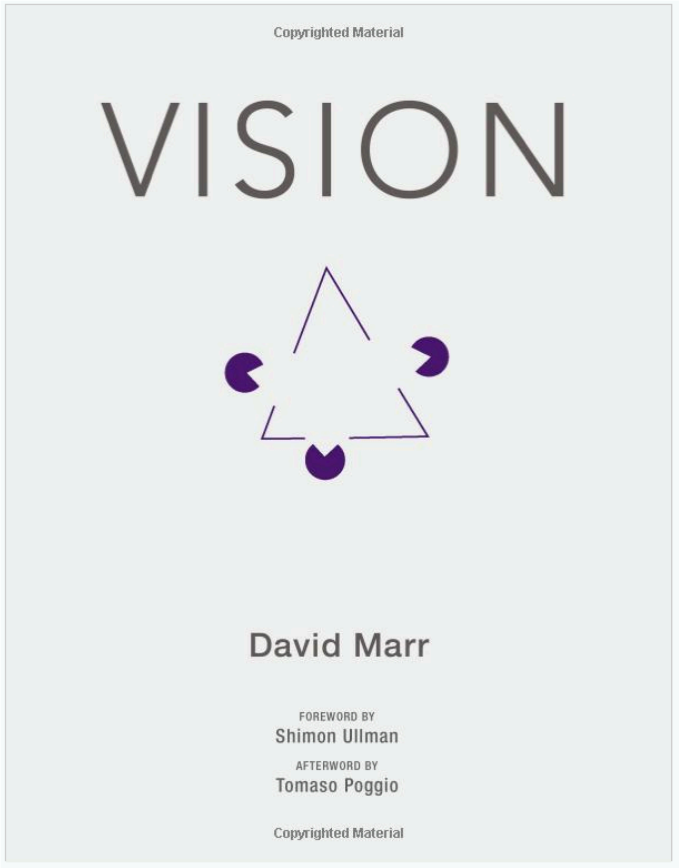
THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".



https://dspace.mit.edu/handle/1721.1/6125



Copyrighted Material

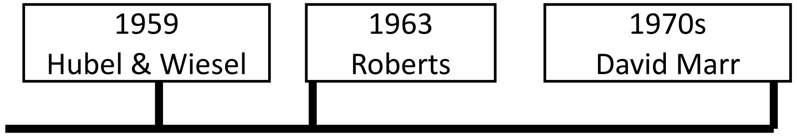
VISION

David Marr

FOREWORD BY
Shimon Ullman

AFTERWORD BY
Tomaso Poggio

Copyrighted Material

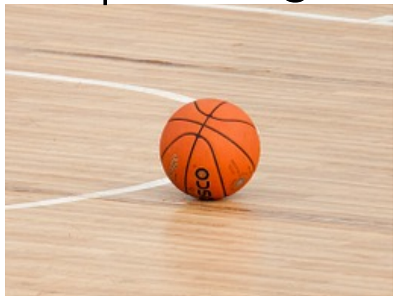


1959
Hubel & Wiesel

1963
Roberts

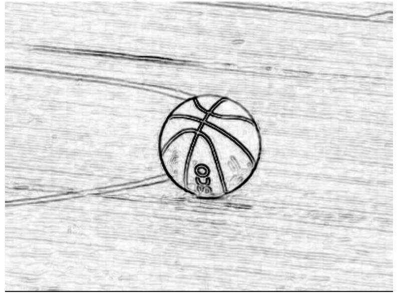
1970s
David Marr

Input image

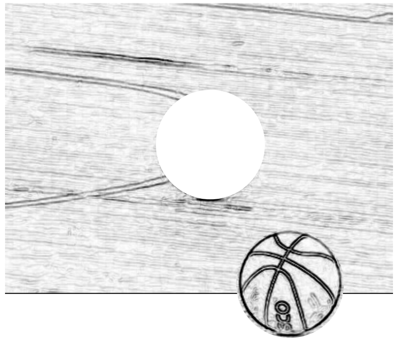


This image is CC0 1.0 public domain

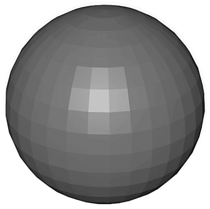
Edge image



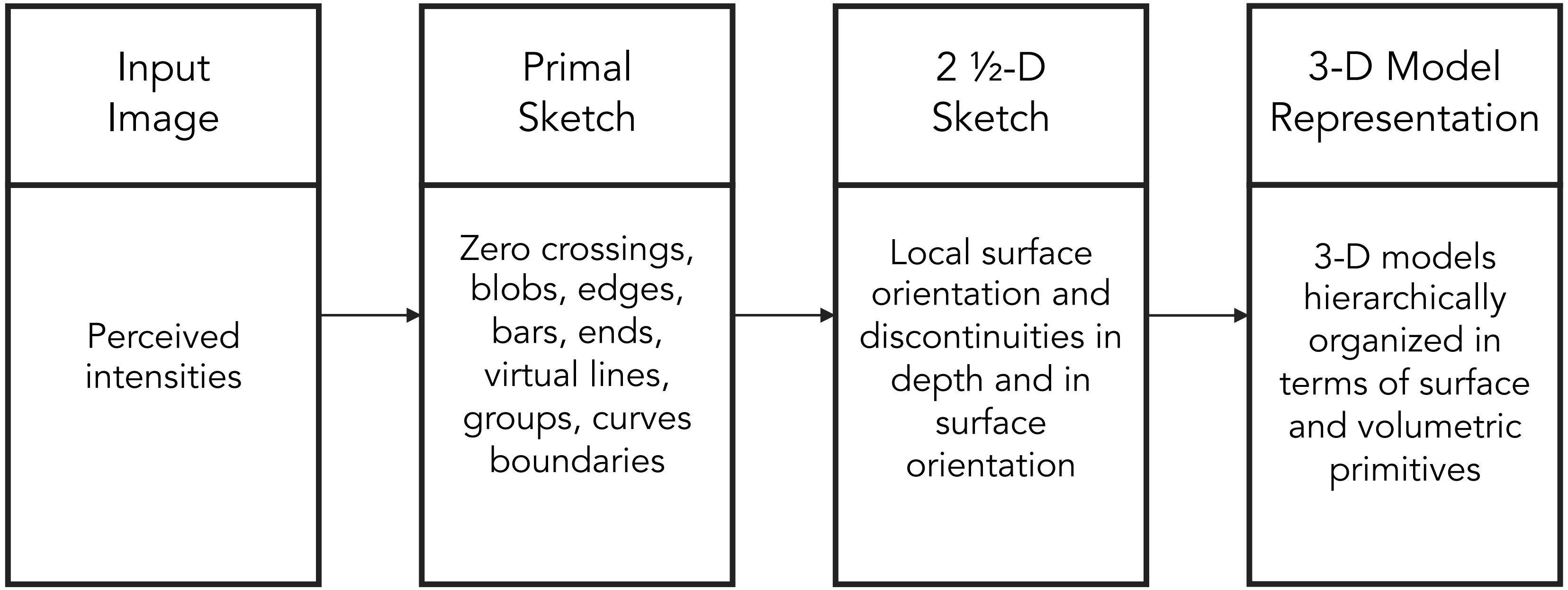
2 ½-D sketch



3-D model



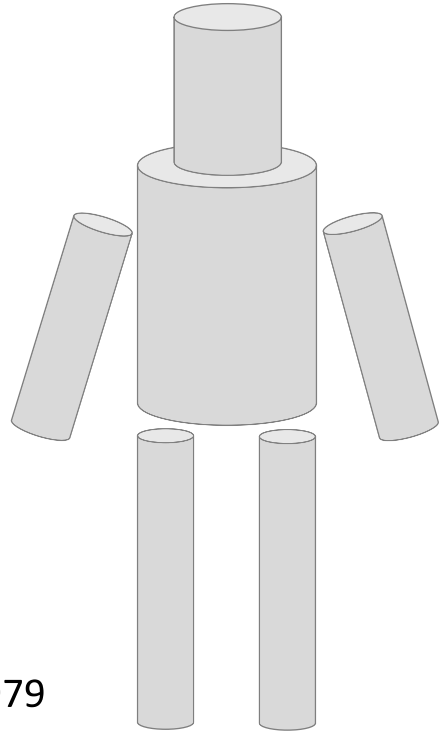
This image is CC0 1.0 public domain



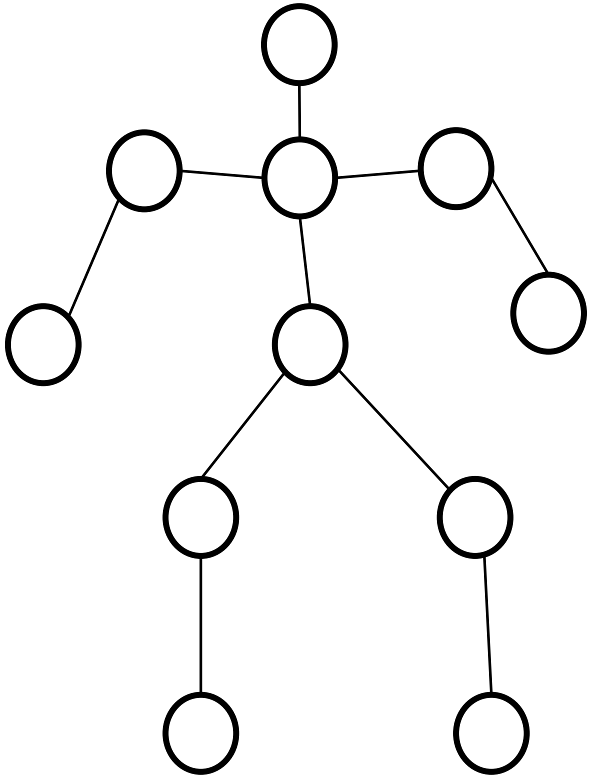
Input Image	Primal Sketch	2 ½-D Sketch	3-D Model Representation
Perceived intensities	Zero crossings, blobs, edges, bars, ends, virtual lines, groups, curves boundaries	Local surface orientation and discontinuities in depth and in surface orientation	3-D models hierarchically organized in terms of surface and volumetric primitives

Stages of Visual Representation, David Marr, 1970s

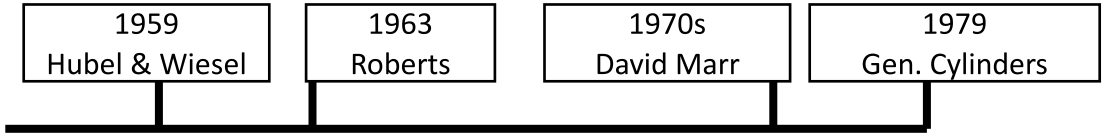
Recognition via Parts (1970s)



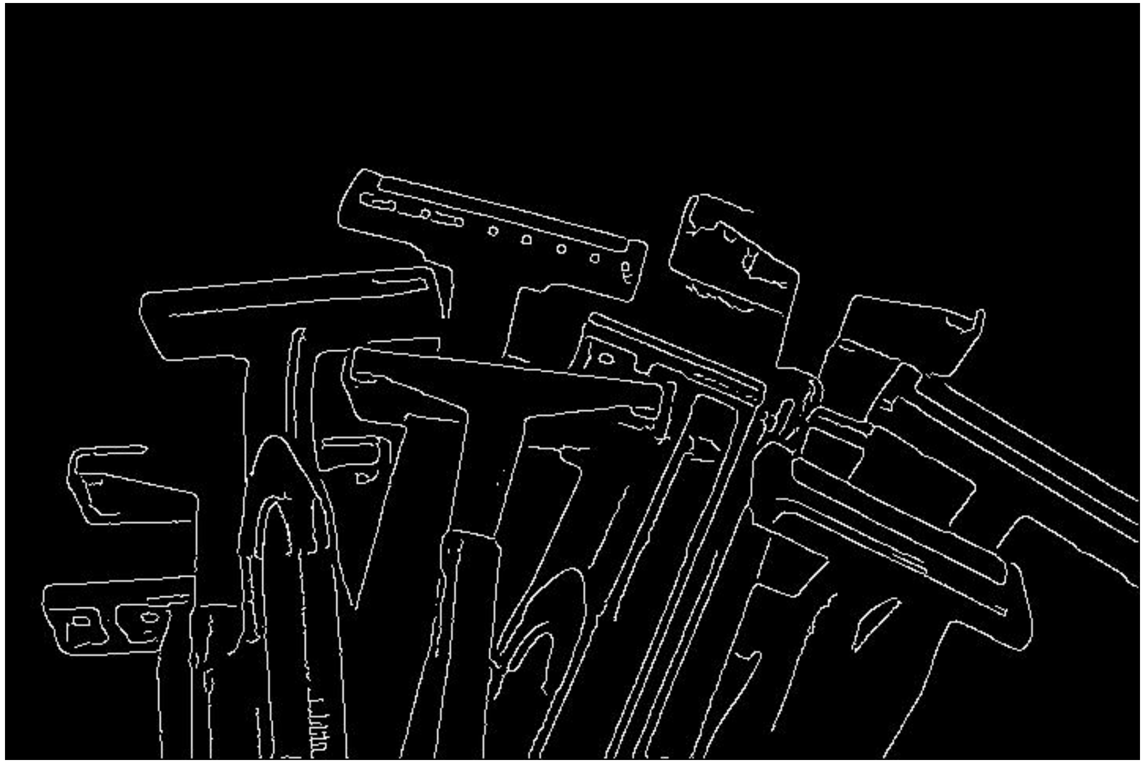
Generalized Cylinders,
Brooks and Binford, 1979

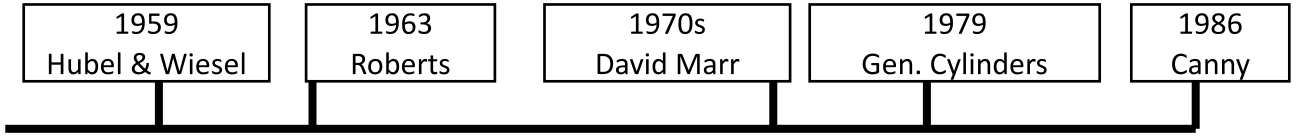


Pictorial Structures,
Fischler and Elshlager, 1973



Recognition via Edge Detection (1980s)





1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

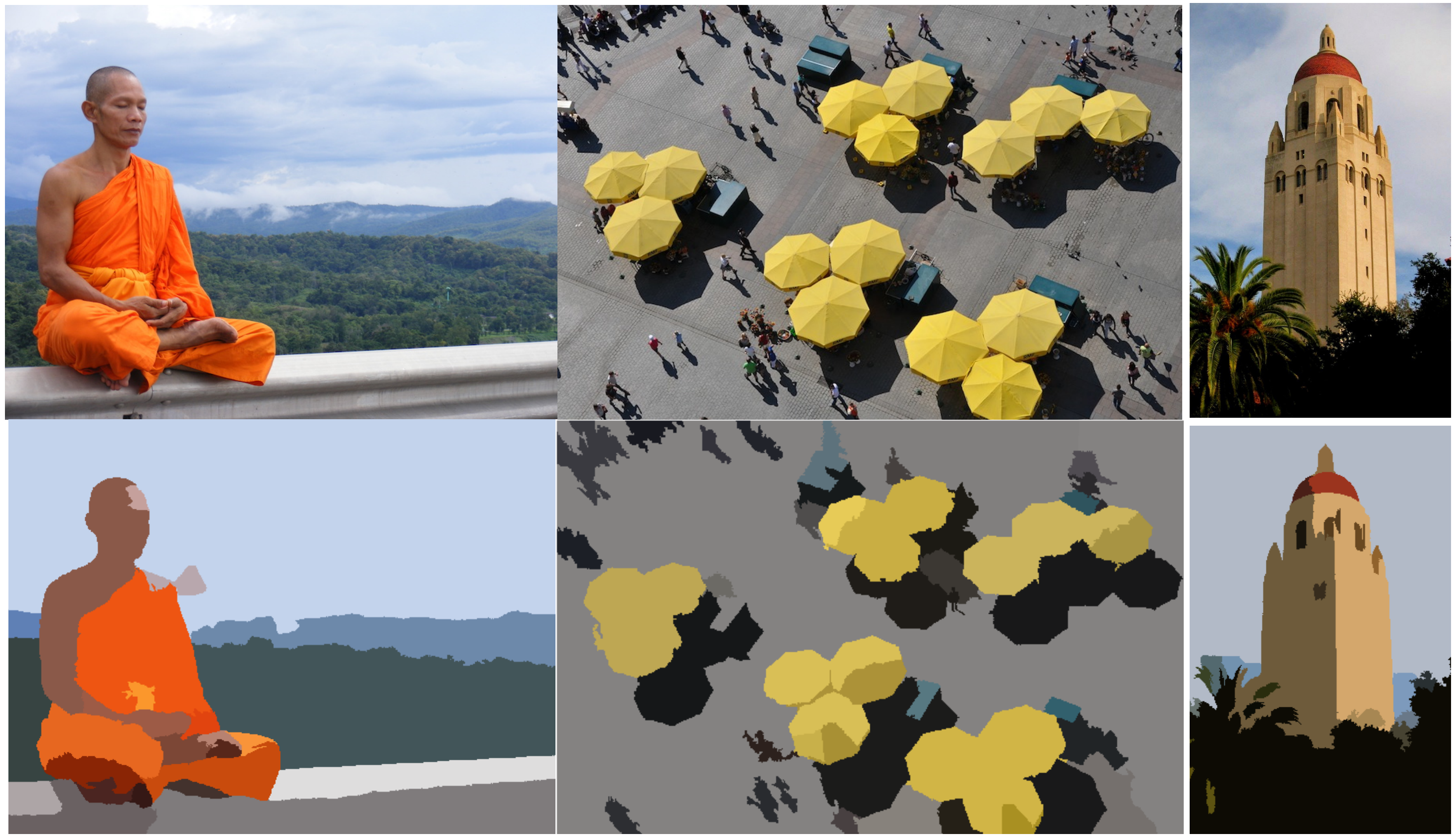
1979
Gen. Cylinders

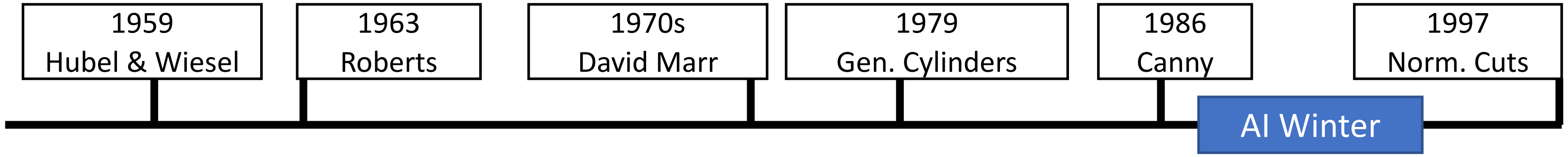
1986
Canny

John Canny, 1986
David Lowe, 1987

Image is CC0 1.0 public domain

Recognition via Grouping (1990s)





Normalized Cuts, Shi and Malik, 1997

Recognition via Matching (2000s)

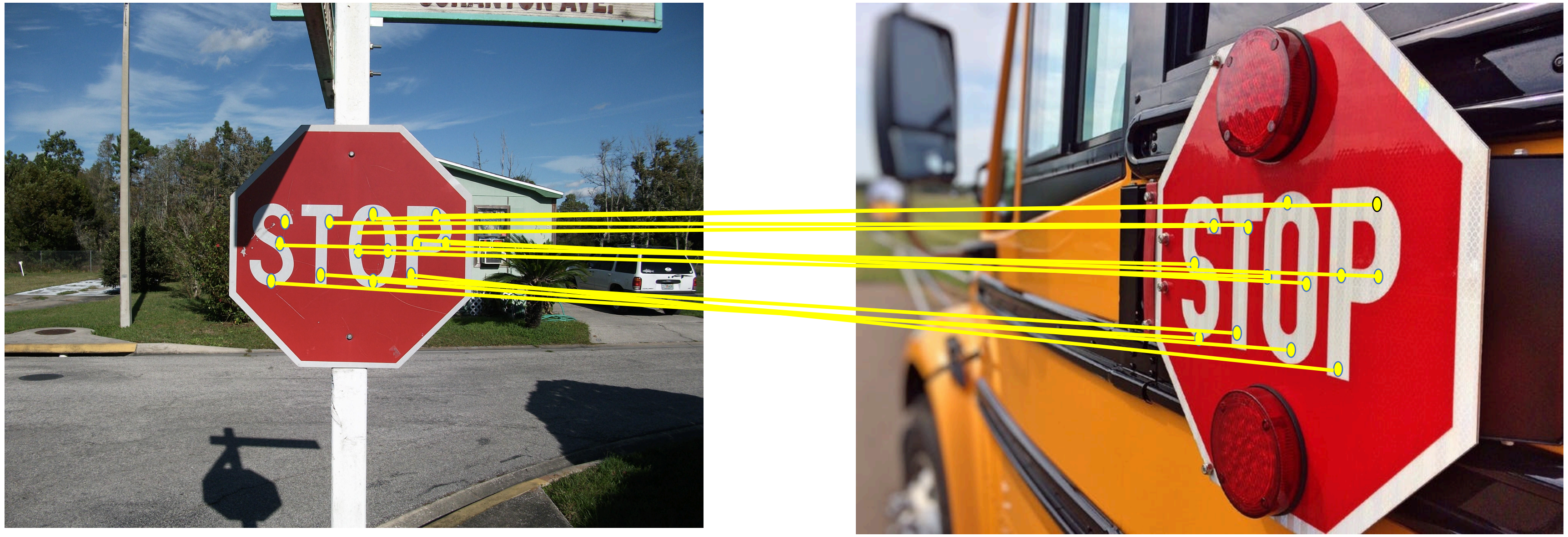
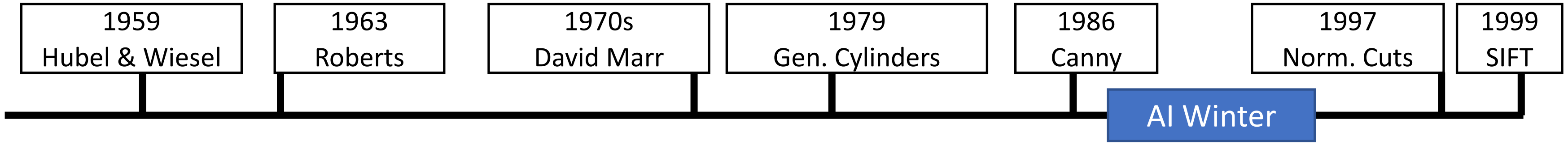


Image is public domain

Image is public domain



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

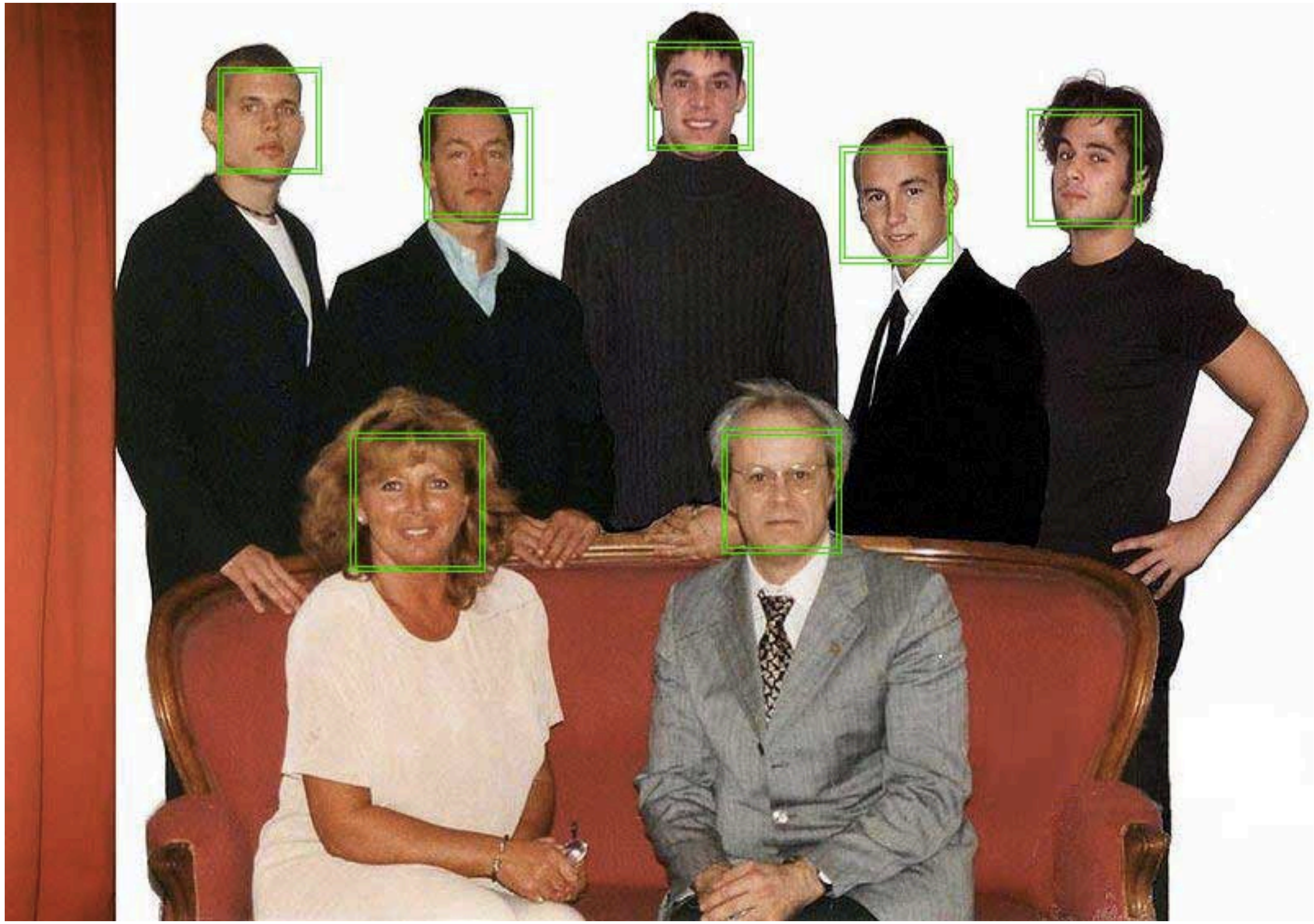
AI Winter

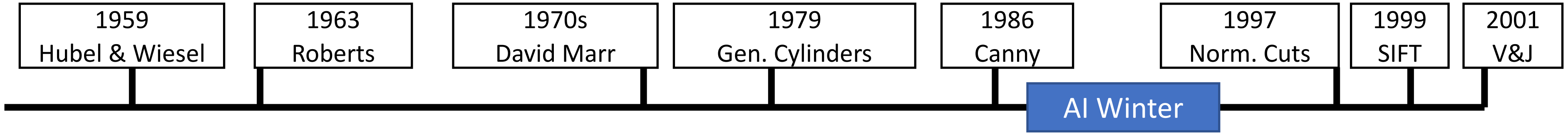
SIFT, David Lowe, 1999

Face Detection

Viola and Jones, 2001

One of the first successful applications of machine learning to vision





1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

AI Winter

1997
Norm. Cuts

1999
SIFT

2001
V&J

PASCAL Visual Object Challenge

Image is CC0 1.0 public domain

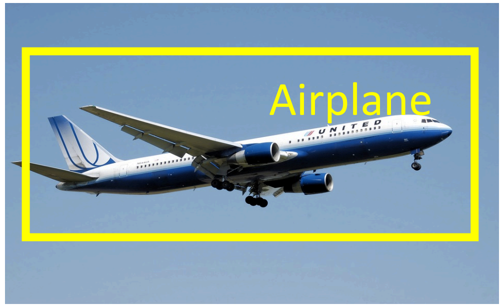
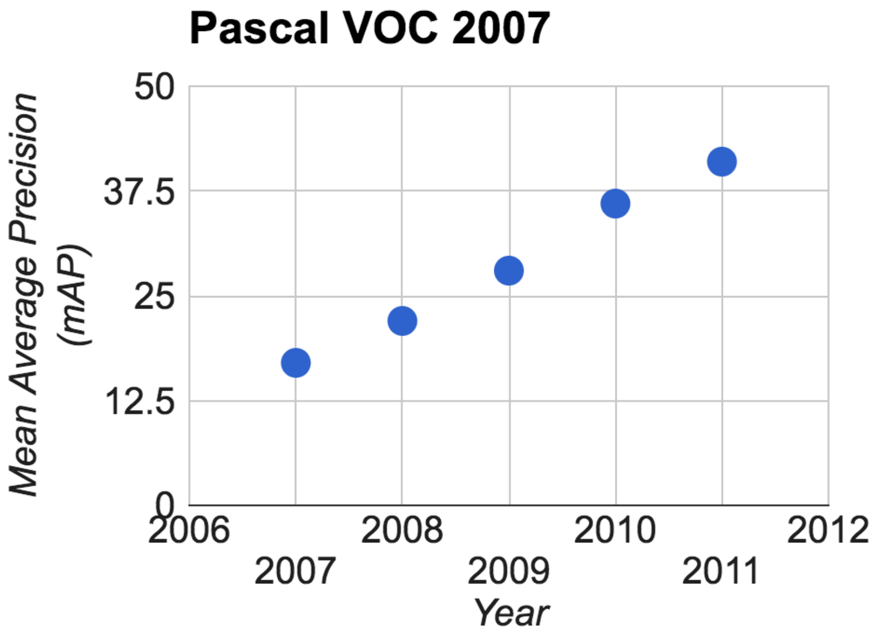
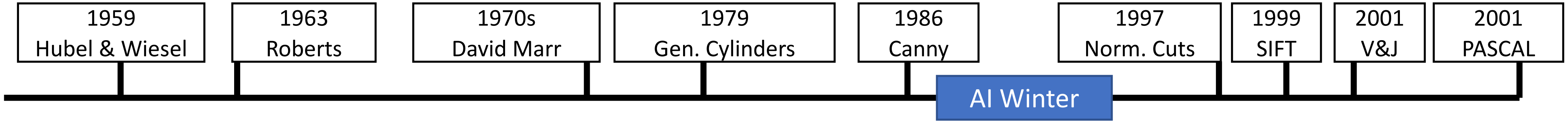


Image is CC0 1.0 public domain




1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

AI Winter

1997 Norm. Cuts

1999 SIFT

2001 V&J

2001 PASCAL

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:

Scale

T-shirt

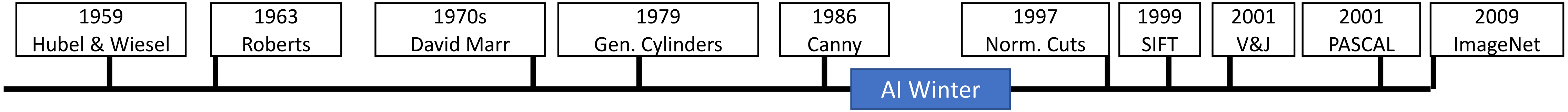
Steel drum

Drumstick

Mud turtle

Deng et al, 2009

Russakovsky et al. IJCV 2015



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

AI Winter

1997
Norm. Cuts

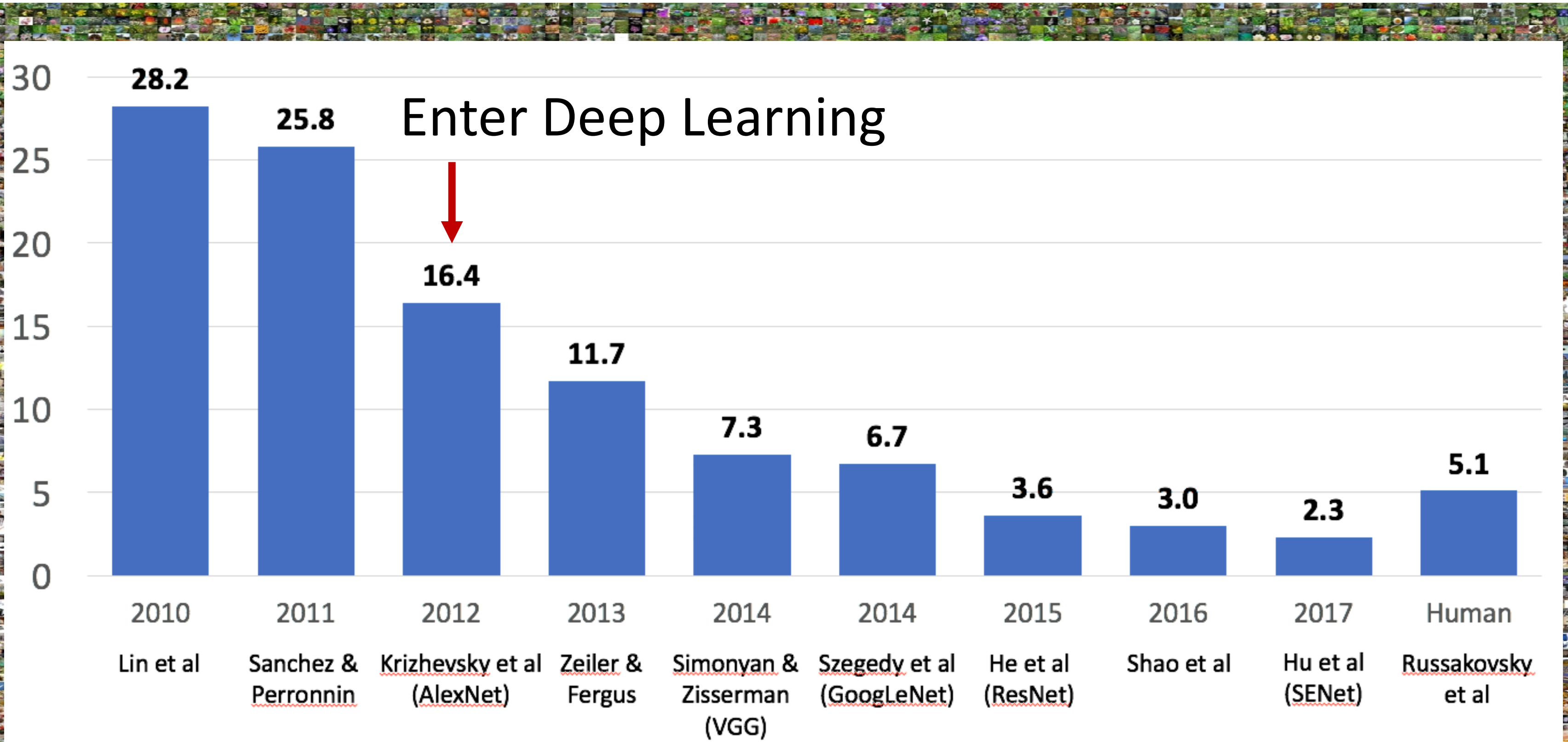
1999
SIFT

2001
V&J

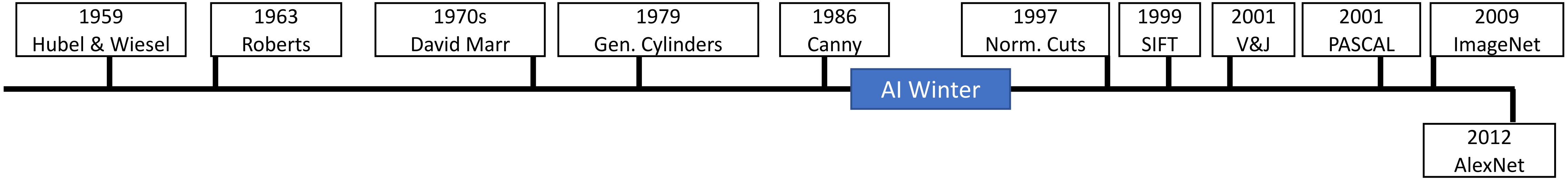
2001
PASCAL

2009
ImageNet

IMAGENET Large Scale Visual Recognition Challenge



Enter Deep Learning



1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

AI Winter

1997 Norm. Cuts

1999 SIFT

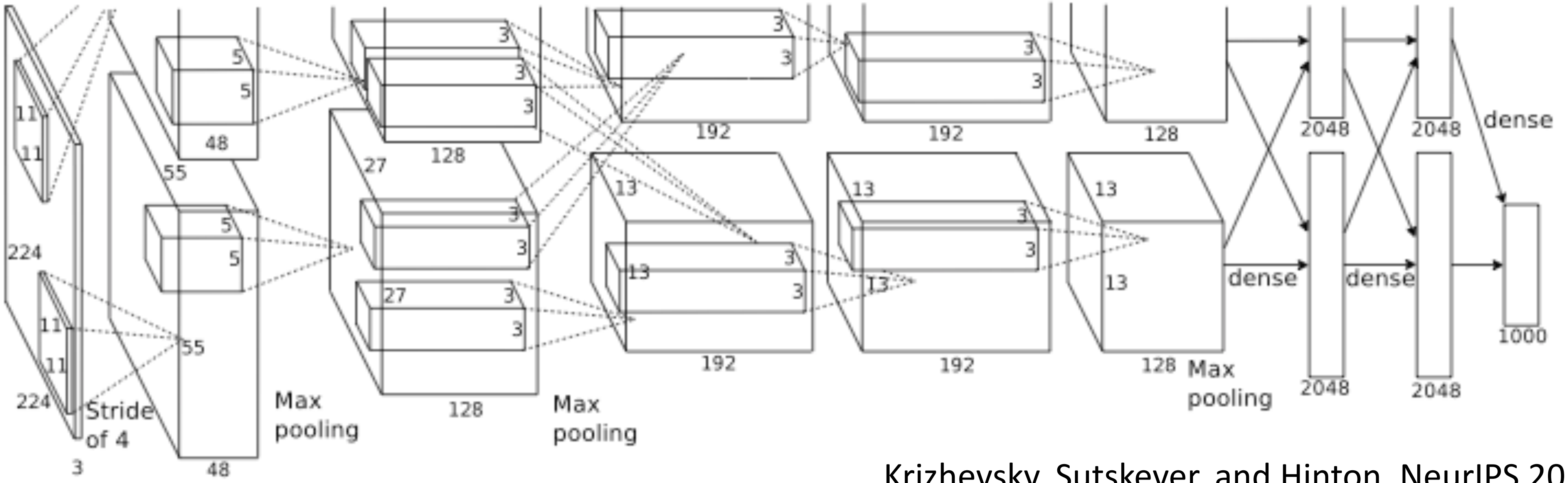
2001 V&J

2001 PASCAL

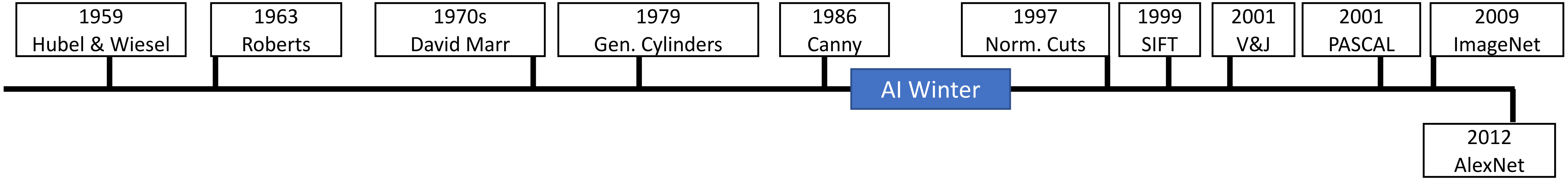
2009 ImageNet

2012 AlexNet

AlexNet: Deep Learning Goes Mainstream



Krizhevsky, Sutskever, and Hinton, NeurIPS 2012



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

AI Winter

1997
Norm. Cuts

1999
SIFT

2001
V&J

2001
PASCAL

2009
ImageNet

2012
AlexNet

Perceptron

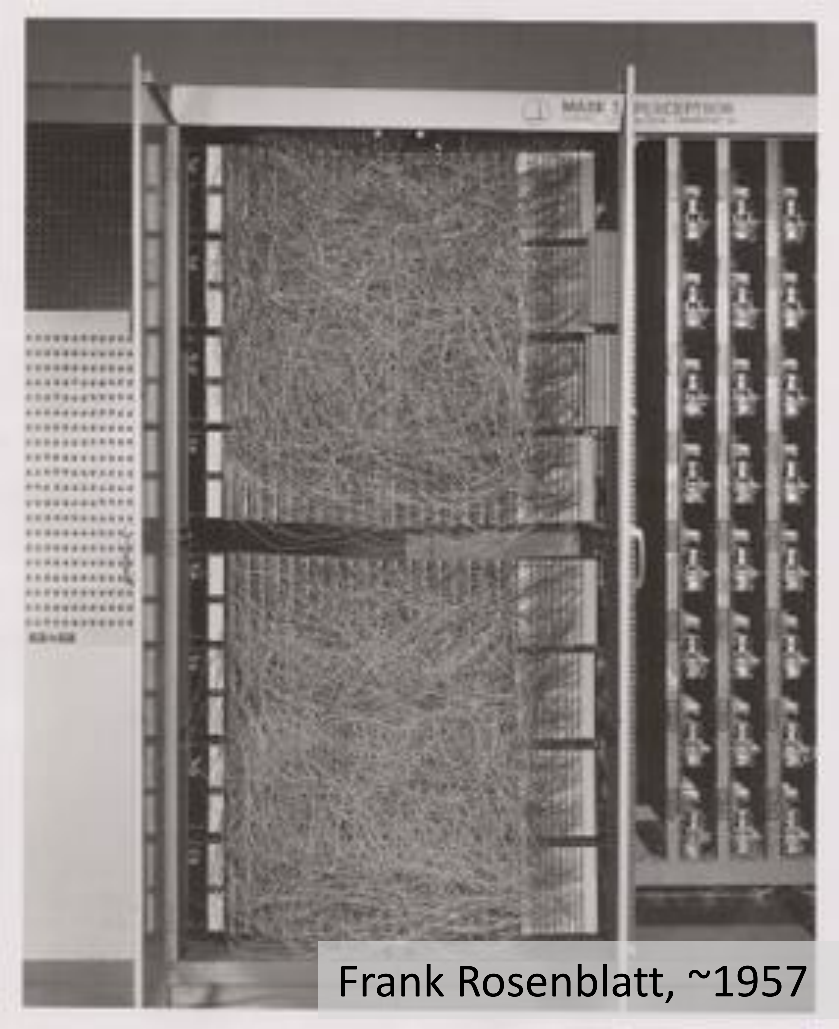
One of the earliest algorithms that could learn from data

Implemented in hardware! Weights stored in potentiometers, updated with electric motors during learning

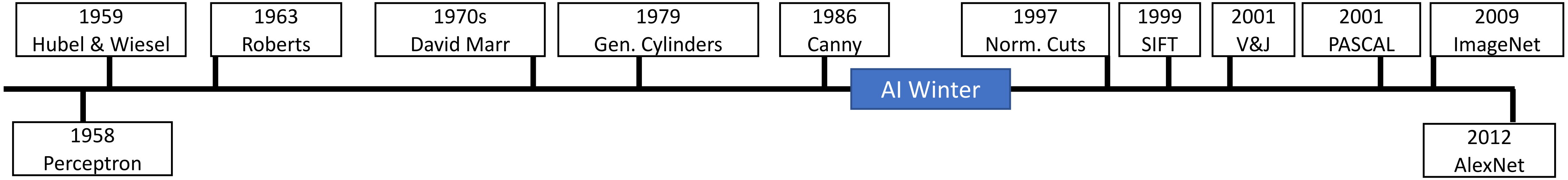
Connected to a camera that used 20x20 cadmium sulfide photocells to make a 400-pixel image

Could learn to recognize letters of the alphabet

Today we would recognize it as a **linear classifier**



Frank Rosenblatt, ~1957



1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

1999 SIFT

2001 V&J

2001 PASCAL

2009 ImageNet

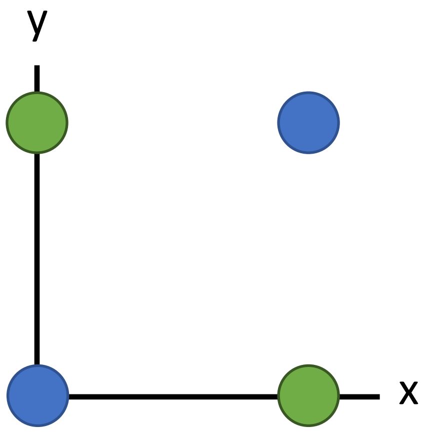
AI Winter

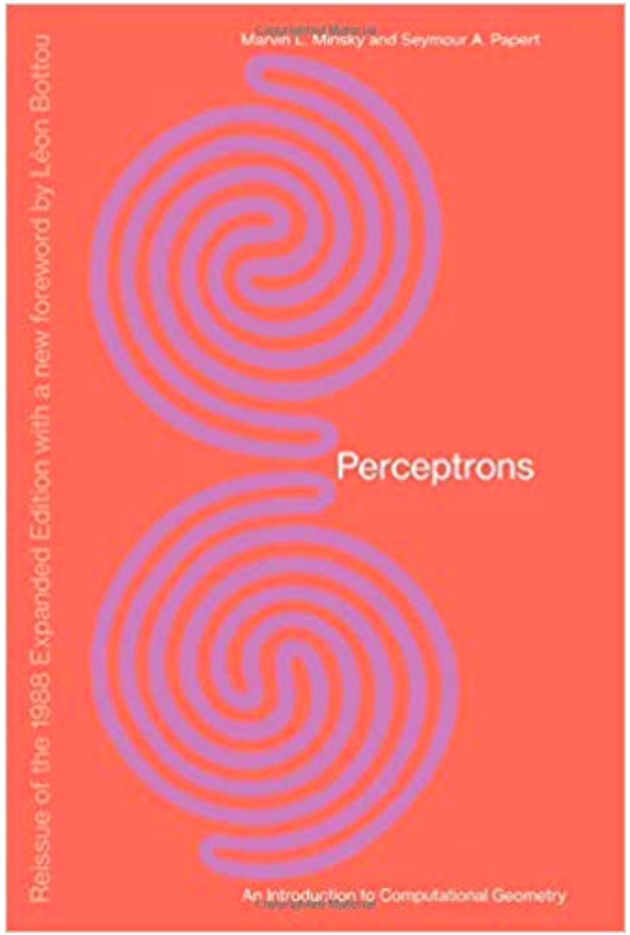
1958 Perceptron

2012 AlexNet

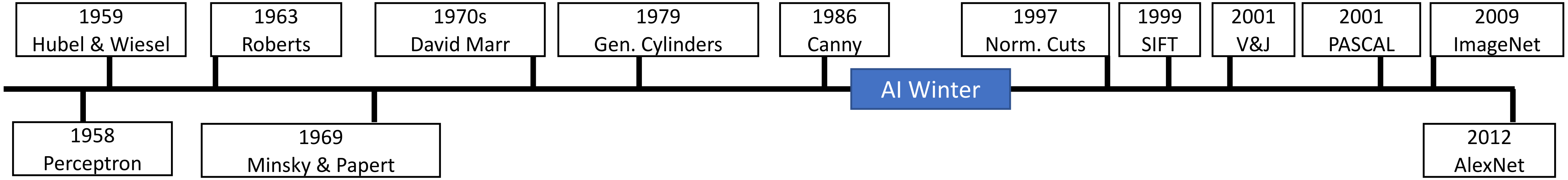
Minsky and Papert, 1969

X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0




Marvin L. Minsky and Seymour A. Papert
Perceptrons
An Introduction to Computational Geometry
Reissue of the 1988 Expanded Edition with a new foreword by Léon Bottou

Showed that Perceptrons could not learn the XOR function
Caused a lot of disillusionment in the field

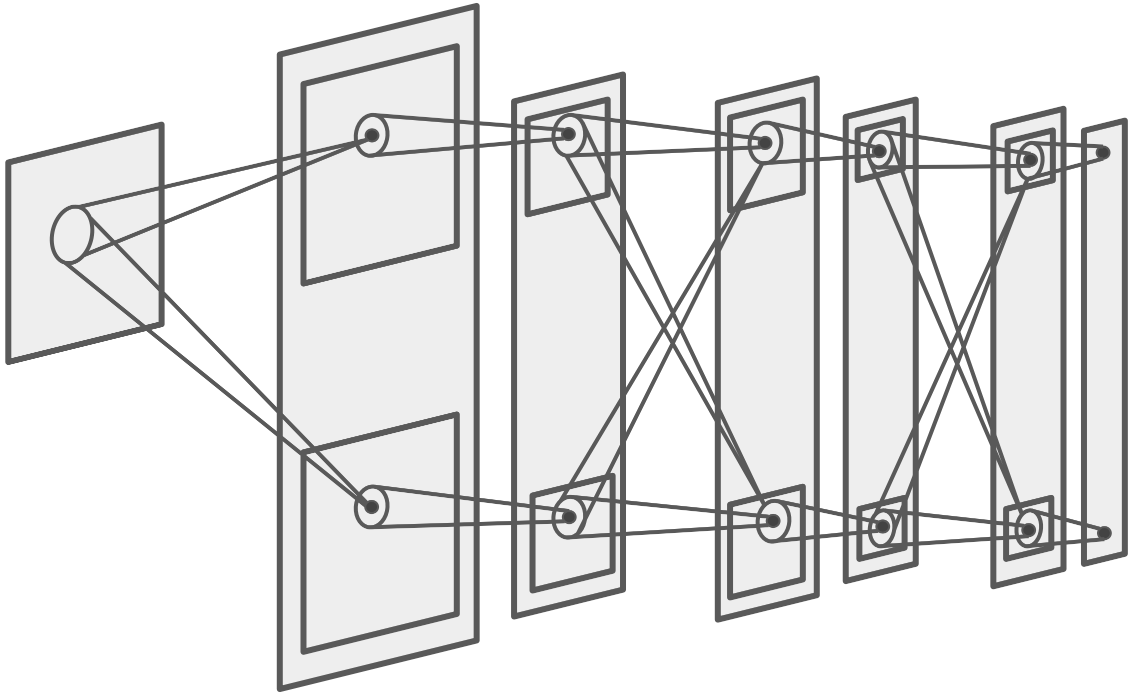


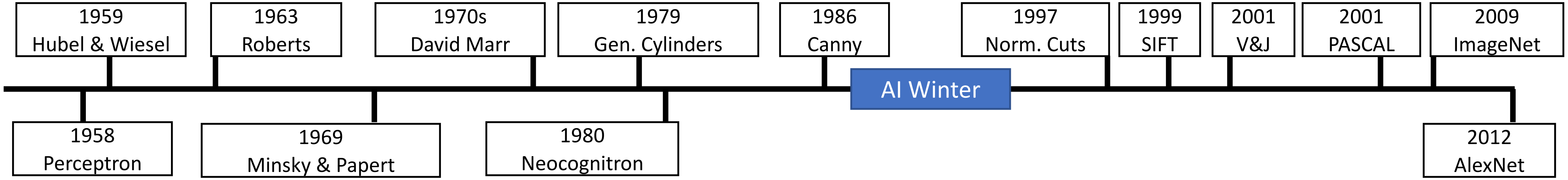
Neocognitron: Fukushima, 1980

Computational model the visual system, directly inspired by Hubel and Wiesel's hierarchy of complex and simple cells

Interleaved simple cells (convolution) and complex cells (pooling)

No practical training algorithm





1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

AI Winter

1997
Norm. Cuts

1999
SIFT

2001
V&J

2001
PASCAL

2009
ImageNet

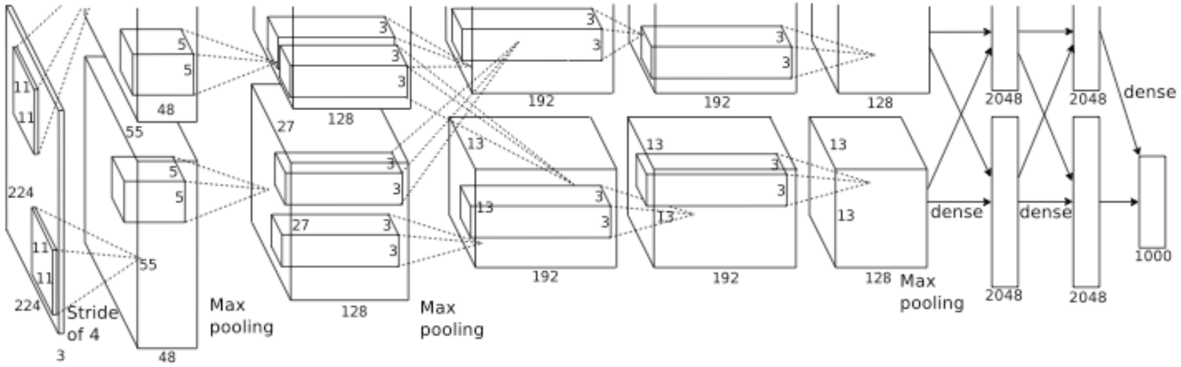
1958
Perceptron

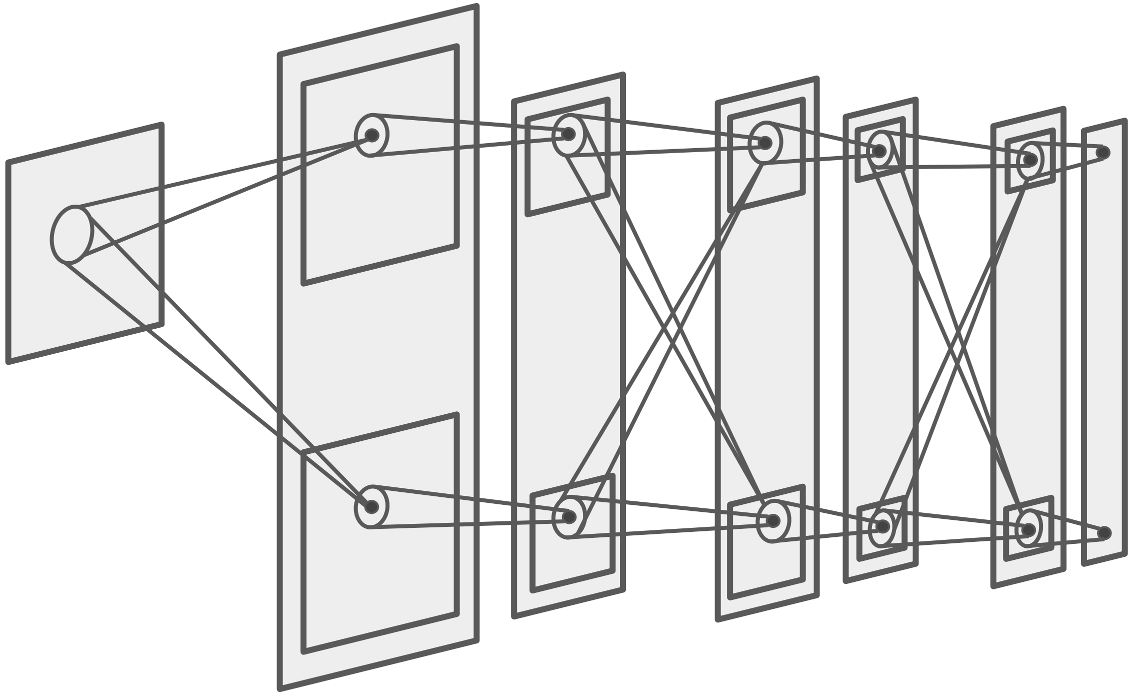
1969
Minsky & Papert

1980
Neocognitron

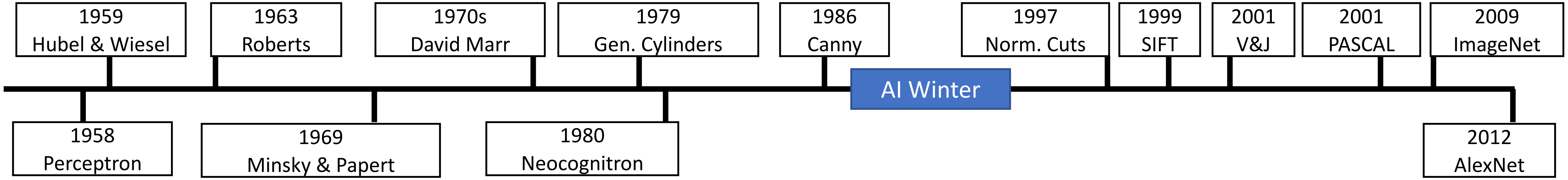
2012
AlexNet

Neocognitron: Fukushima, 1980





Looks a lot like AlexNet
more than 32 years later!



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

2001
PASCAL

2009
ImageNet

AI Winter

1958
Perceptron

1969
Minsky & Papert

1980
Neocognitron

2012
AlexNet

Backprop: Rumelhart, Hinton, and Williams, 1986

Introduced backpropagation for computing gradients in neural networks

Successfully trained perceptrons with multiple layers

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$

recognizable math

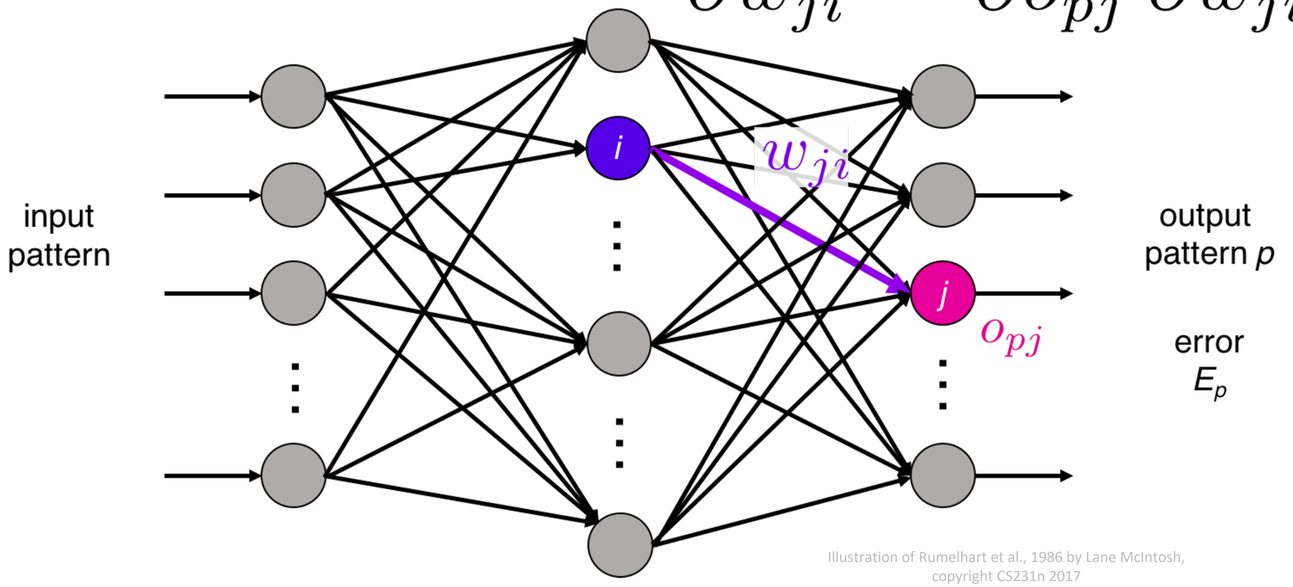
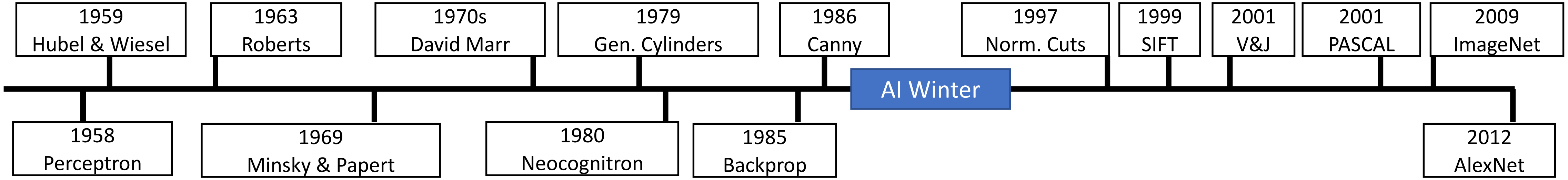


Illustration of Rumelhart et al., 1986 by Lane McIntosh, copyright CS231n 2017



1959 Hubel & Wiesel

1963 Roberts

1970s David Marr

1979 Gen. Cylinders

1986 Canny

1997 Norm. Cuts

1999 SIFT

2001 V&J

2001 PASCAL

2009 ImageNet

AI Winter

1958 Perceptron

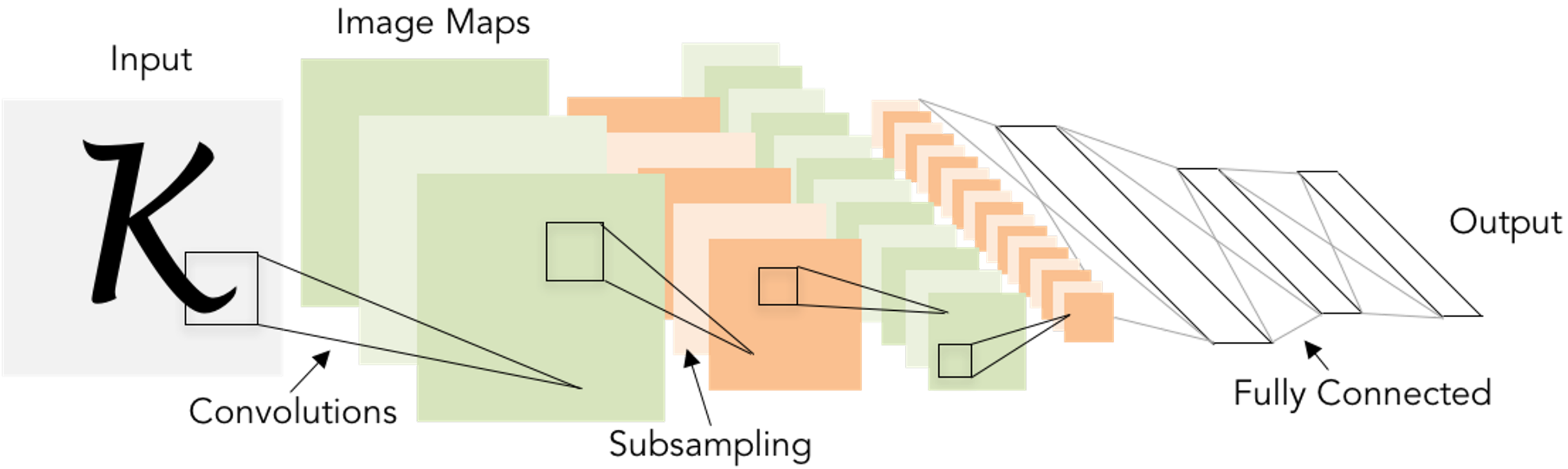
1969 Minsky & Papert

1980 Neocognitron

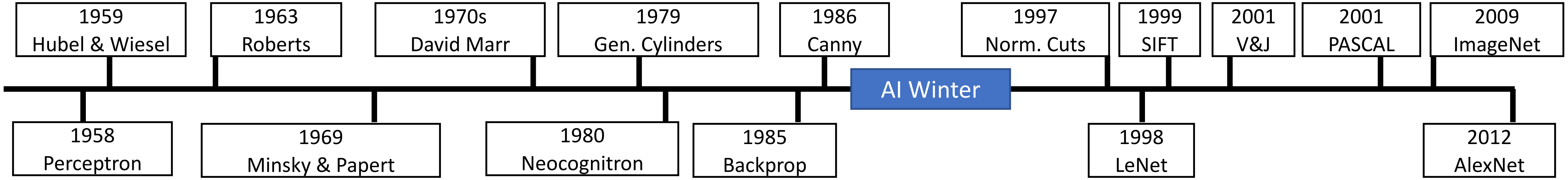
1985 Backprop

2012 AlexNet

Convolutional Networks: LeCun et al, 1998



Applied backprop algorithm to a Neocognitron-like architecture
Learned to recognize handwritten digits
Was deployed in a commercial system by NEC, processed handwritten checks
Very similar to our modern convolutional networks!



1959
Hubel & Wiesel

1963
Roberts

1970s
David Marr

1979
Gen. Cylinders

1986
Canny

1997
Norm. Cuts

1999
SIFT

2001
V&J

2001
PASCAL

2009
ImageNet

1958
Perceptron

1969
Minsky & Papert

1980
Neocognitron

1985
Backprop

AI Winter

1998
LeNet

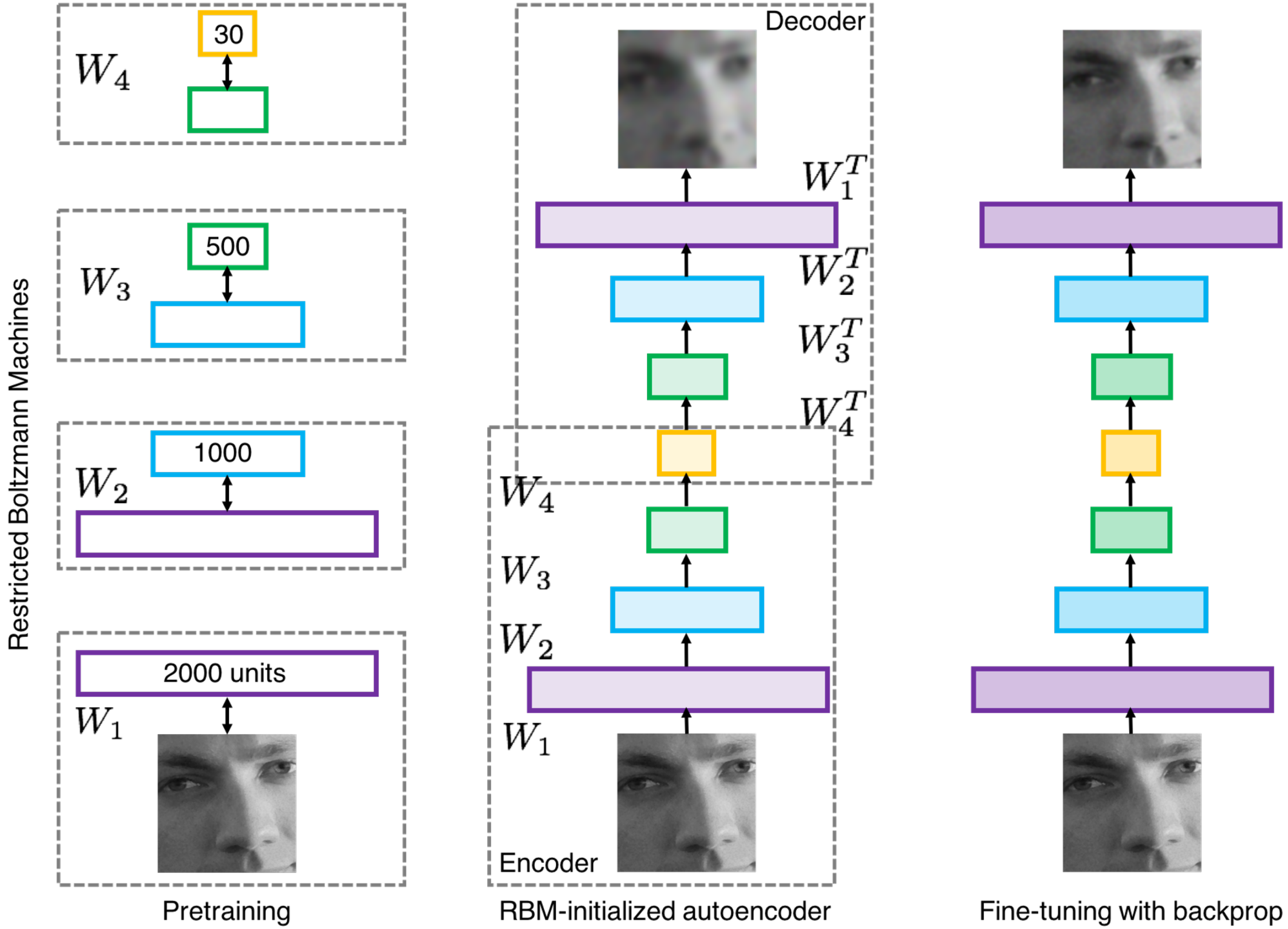
2012
AlexNet

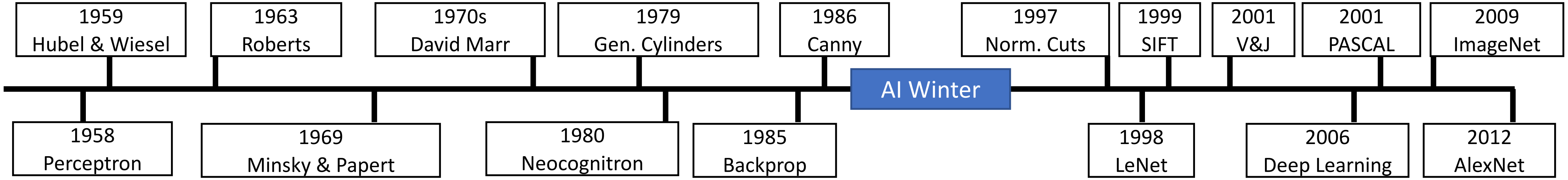
2000s: “Deep Learning”

People tried to train neural networks that were deeper and deeper

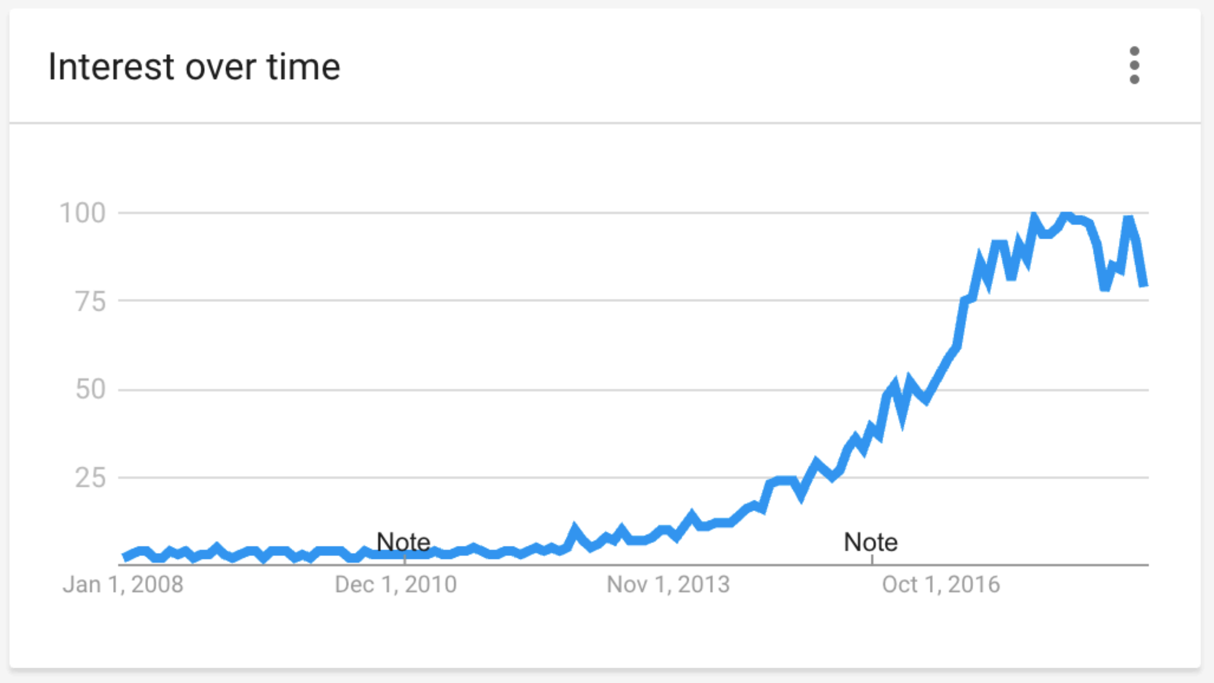
Not a mainstream research topic at this time

Hinton and Salakhutdinov, 2006
Bengio et al, 2007
Lee et al, 2009
Glorot and Bengio, 2010

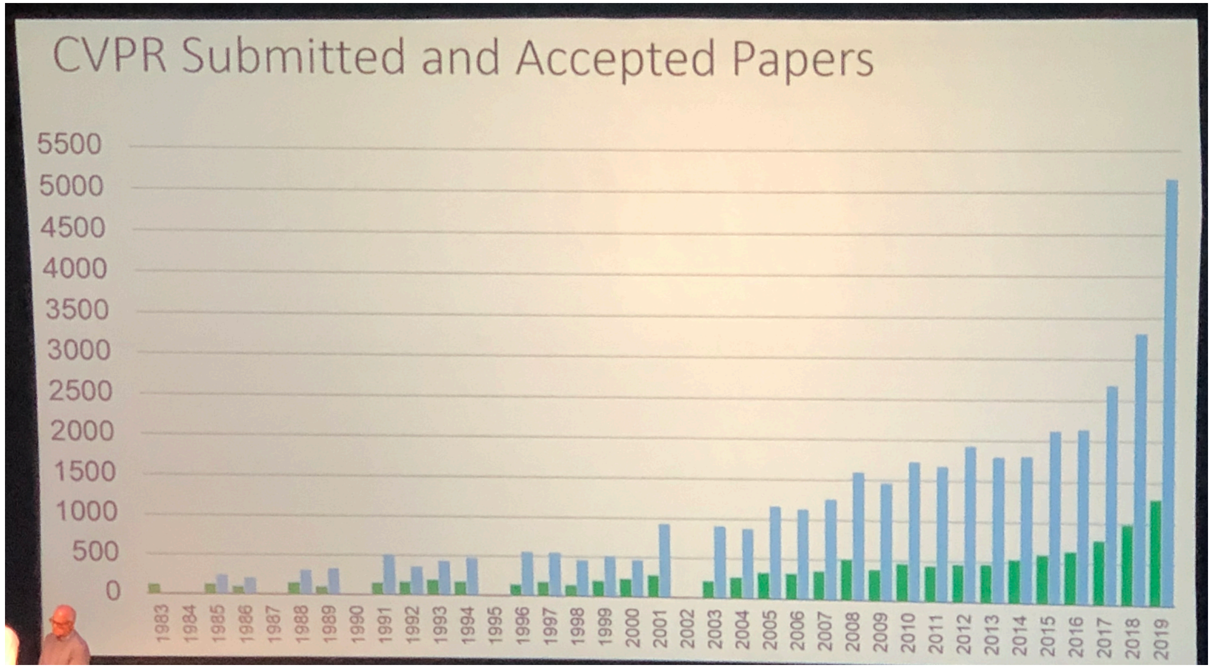




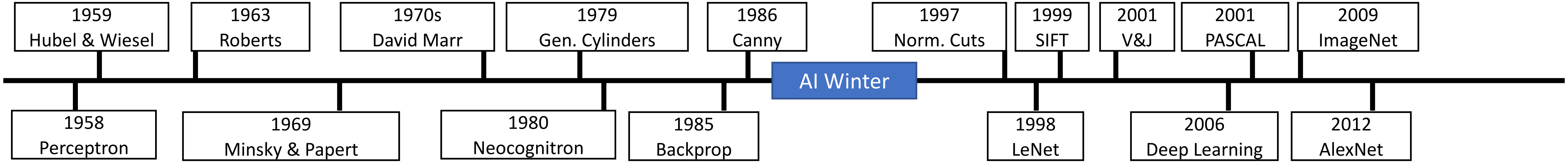
2012 to Present: Deep Learning Explosion



Google Trends: “Deep Learning”



Publications at top Computer Vision conference



1959 Hubel & Wiesel

1958 Perceptron

1963 Roberts

1969 Minsky & Papert

1970s David Marr

1979 Gen. Cylinders

1980 Neocognitron

1985 Backprop

1986 Canny

AI Winter

1997 Norm. Cuts

1998 LeNet

1999 SIFT

2001 V&J

2001 PASCAL

2006 Deep Learning

2009 ImageNet

2012 AlexNet

2012 to Present: ConvNets are everywhere

Image Classification

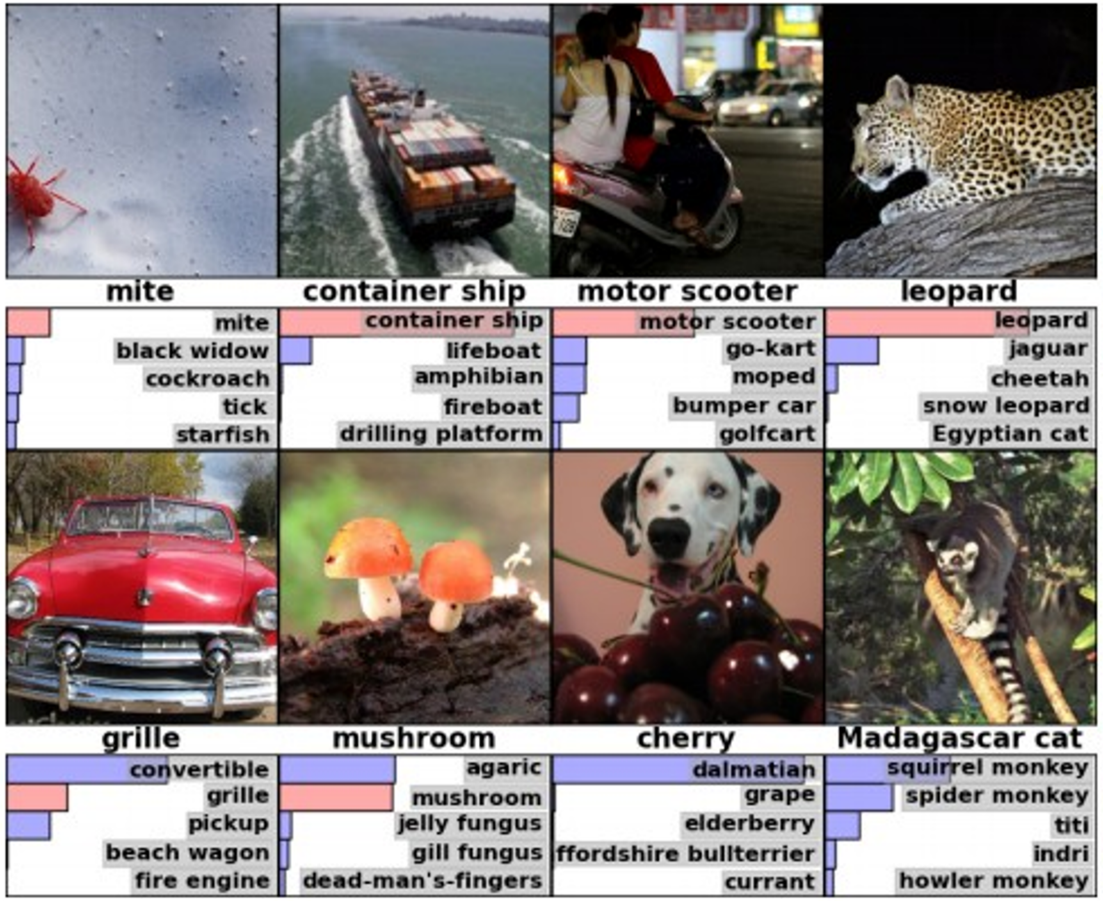
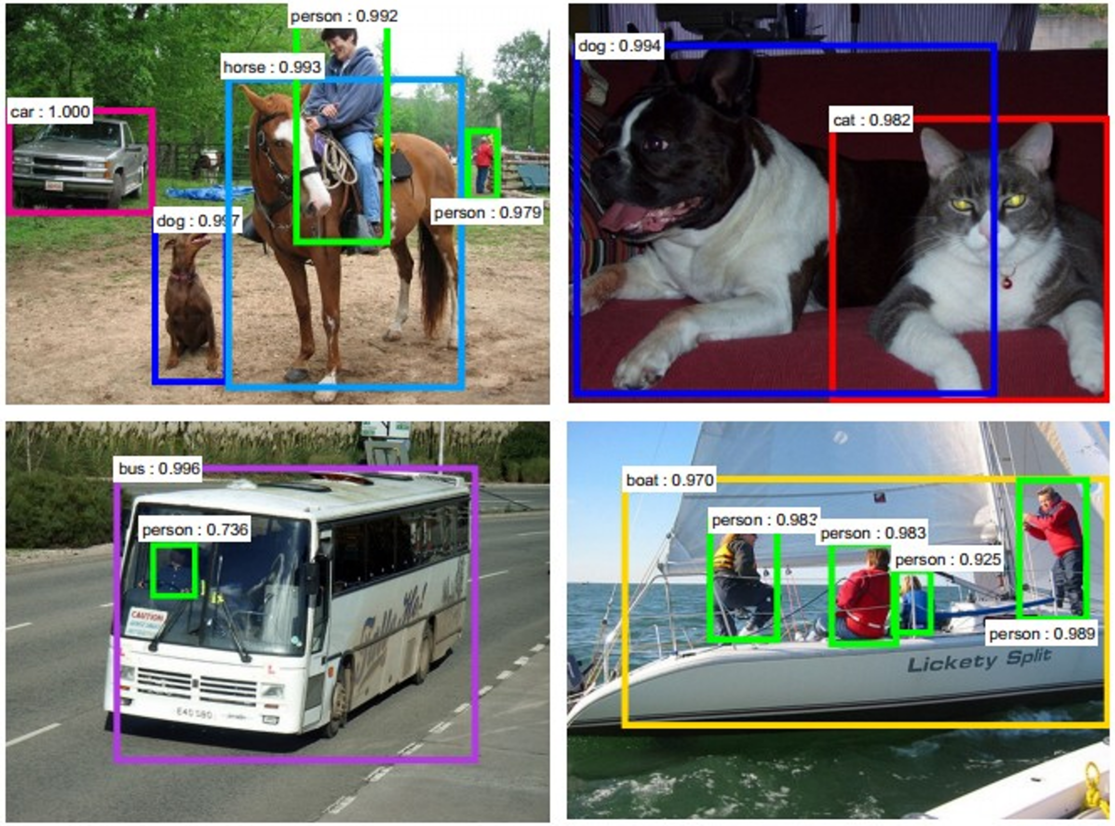


Image Retrieval

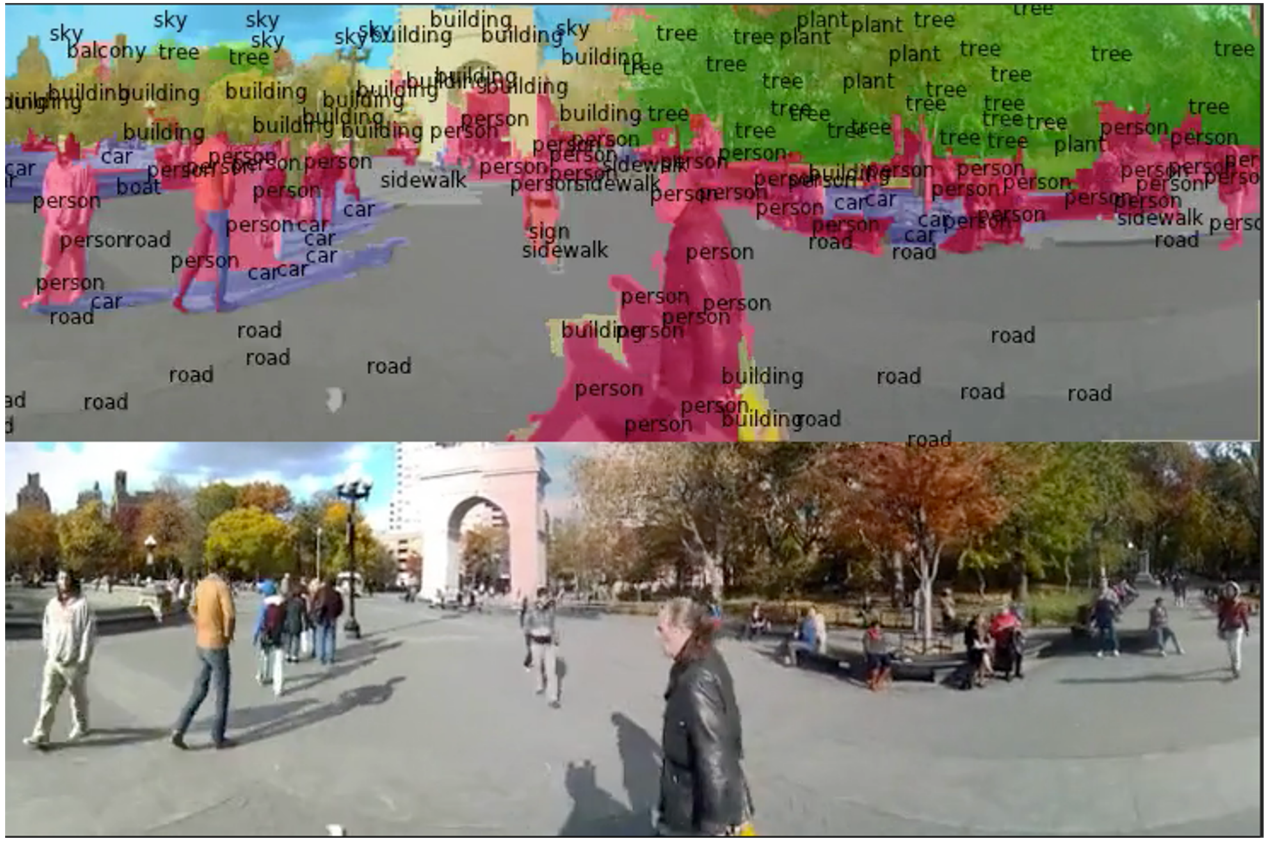
2012 to Present: ConvNets are everywhere

Object Detection



Ren, He, Girshick, and Sun, 2015

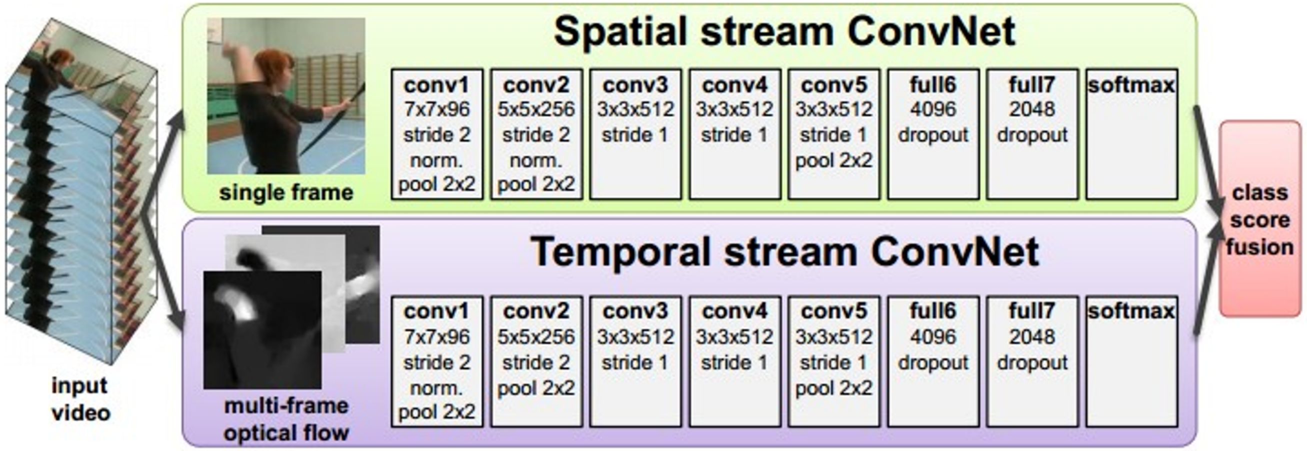
Image Segmentation



Fabaret et al, 2012

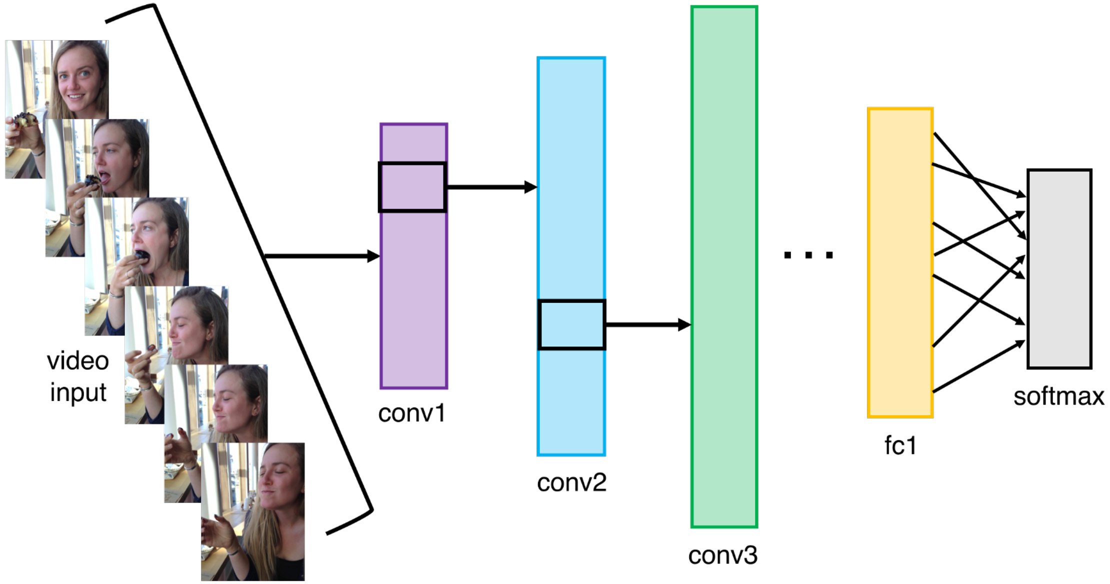
2012 to Present: ConvNets are everywhere

Video Classification



Simonyan et al, 2014

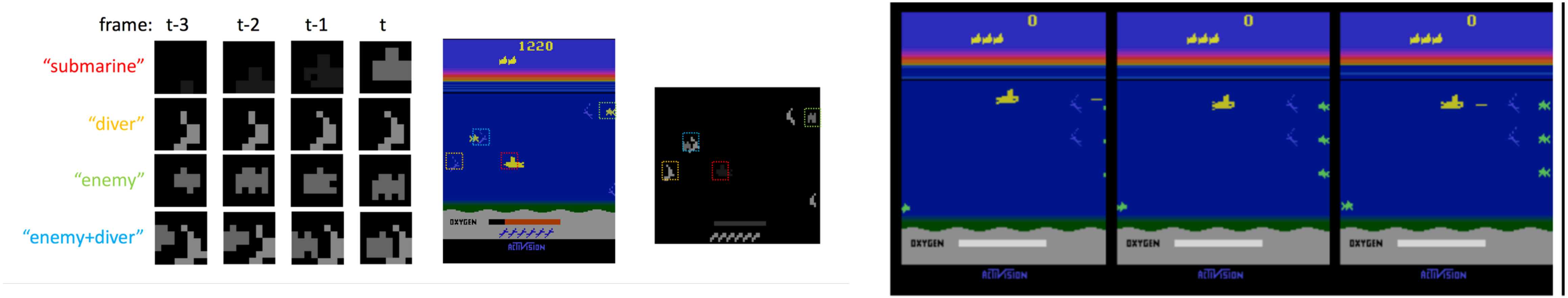
Activity Recognition



2012 to Present: ConvNets are everywhere

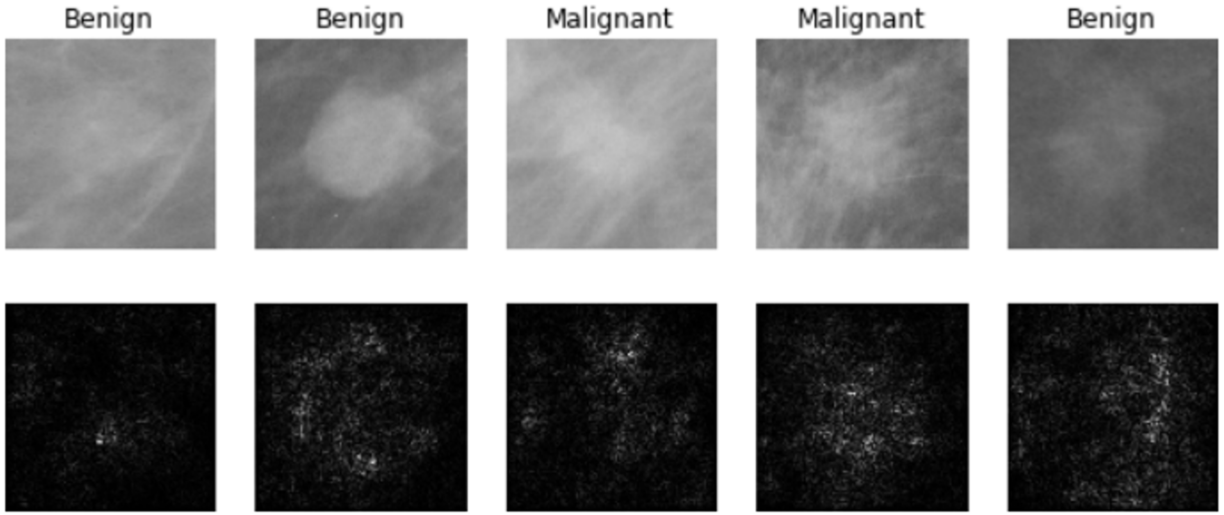
Pose Recognition (Toshev and Szegedy, 2014)

Playing Atari games (Guo et al, 2014)



2012 to Present: ConvNets are everywhere

Medical Imaging



Levy et al, 2016 Figure reproduced with permission

Galaxy Classification

Dieleman et al, 2014

From left to right: public domain by NASA, usage permitted by ESA/Hubble, public domain by NASA, and public domain.

Whale recognition

Kaggle Challenge

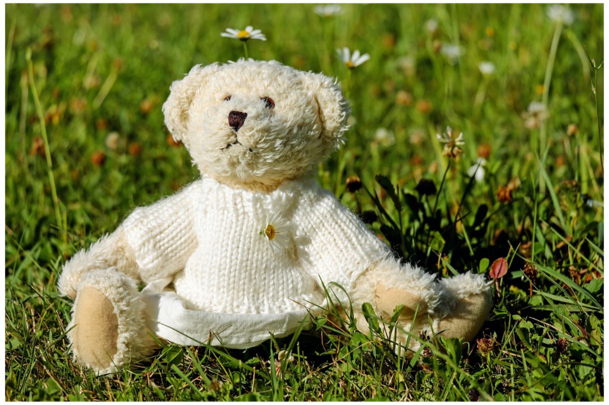
This image by Christin Khan is in the public domain and originally came from the U.S. NOAA.

2012 to Present: ConvNets are everywhere

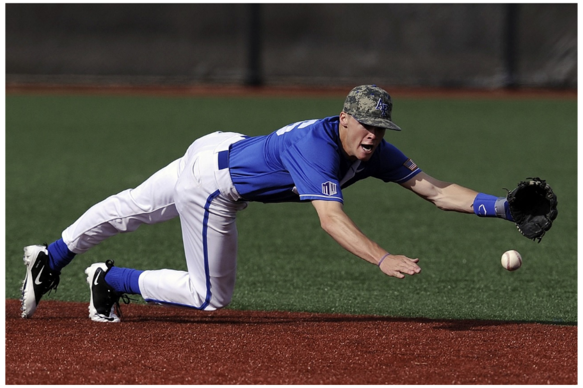
Image Captioning

Vinyals et al, 2015

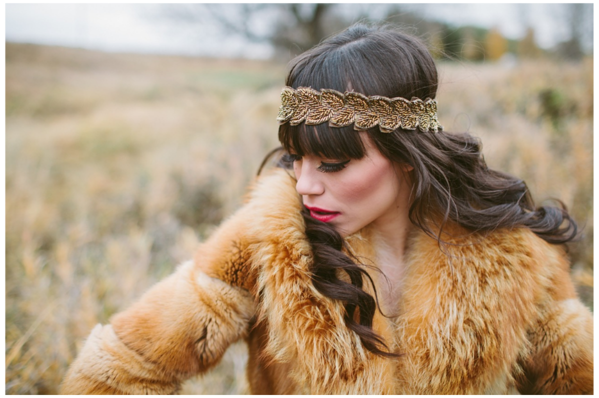
Karpathy and Fei-Fei, 2015



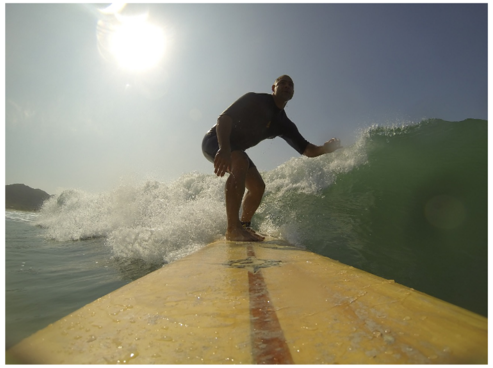
A white teddy bear sitting in the grass



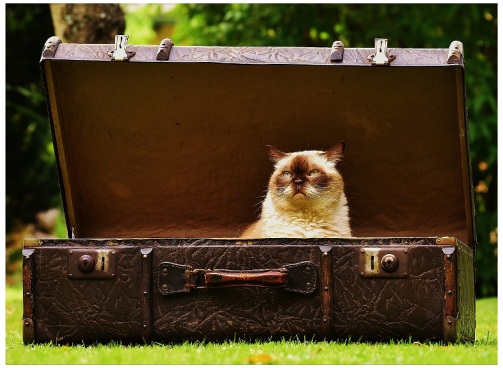
A man in a baseball uniform throwing a ball



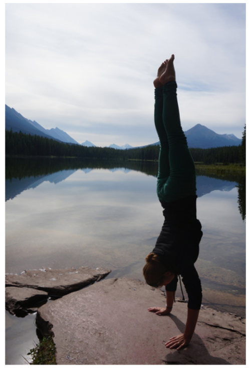
A woman is holding a cat in her hand



A man riding a wave on top of a surfboard



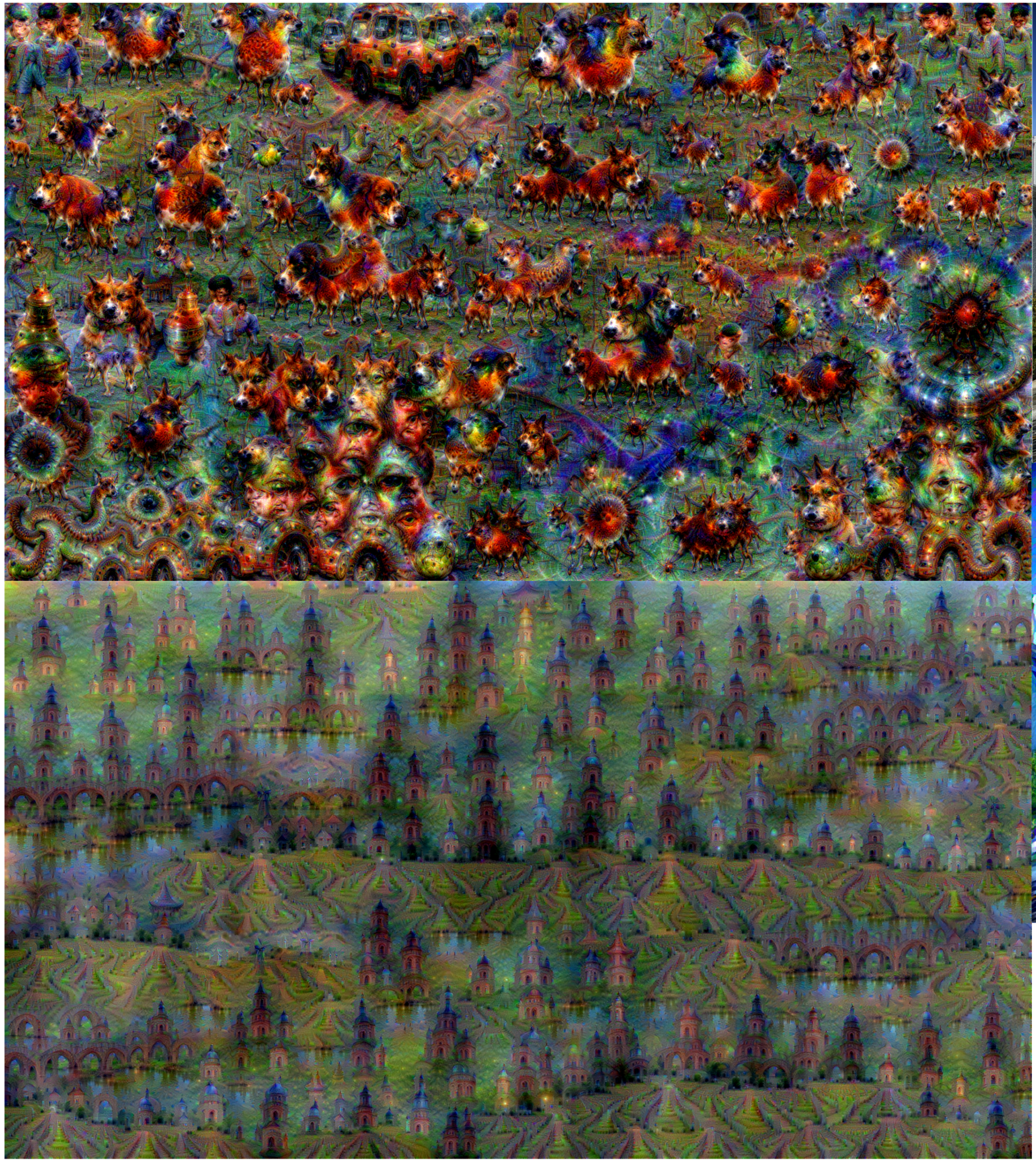
A cat sitting on a suitcase on the floor



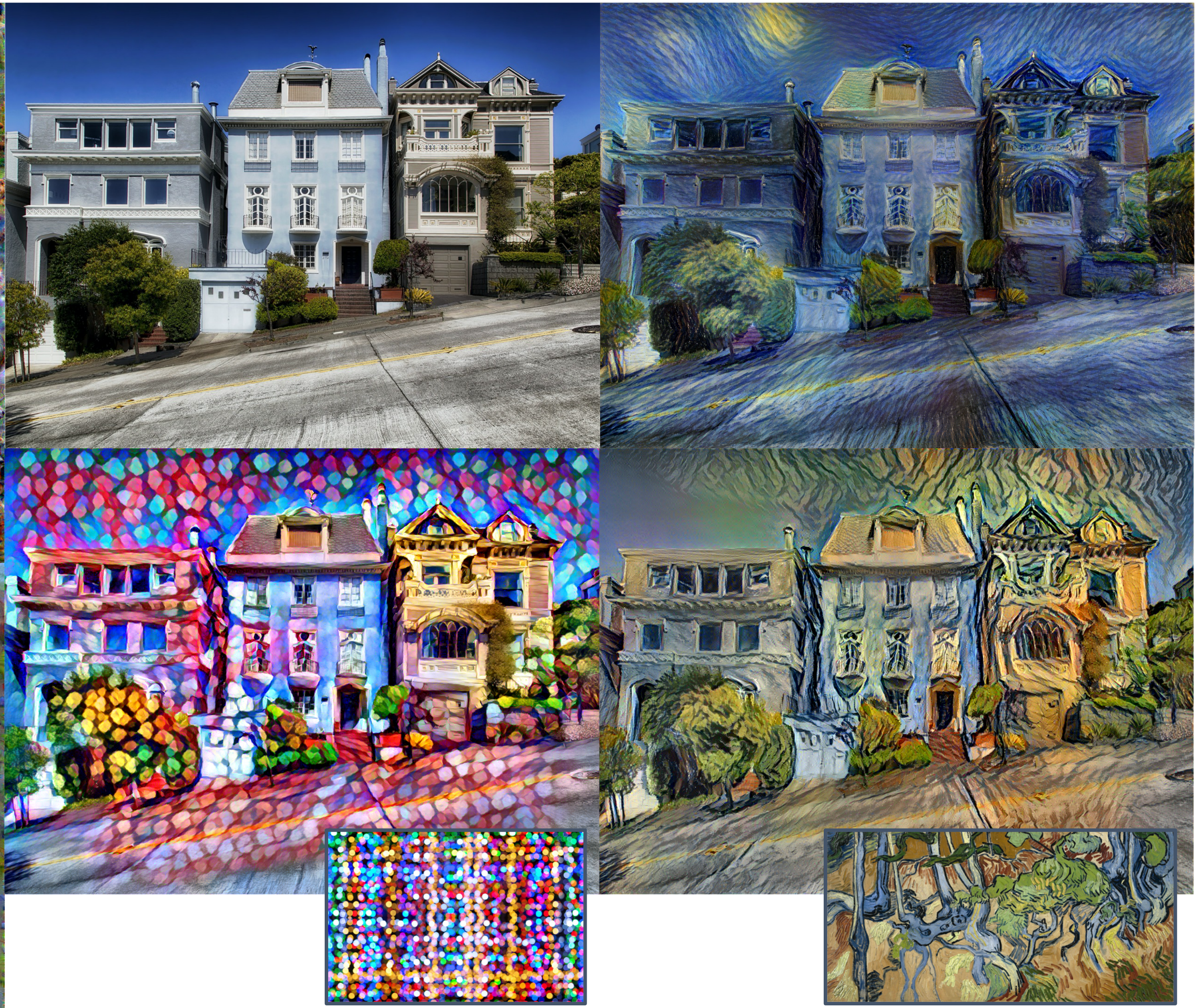
A woman standing on a beach holding a surfboard

https://pixabay.com/en/luggage-antique-cat-1643010/
https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-1623436/
https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/
https://pixabay.com/en/woman-female-model-portrait-adult-983967/
https://pixabay.com/en/handstand-lake-meditation-496008/
https://pixabay.com/en/baseball-player-shortstop-infield-1045263/

Captions generated by Justin Johnson using Neuraltalk2



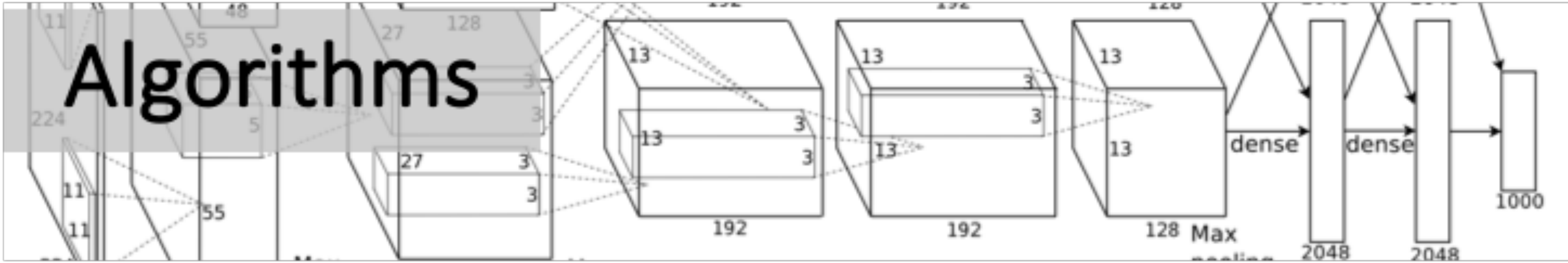
Figures copyright Justin Johnson, 2015. Reproduced with permission. Generated using the Inceptionism approach from a blog post by Google Research.



Original image is CC0 public domain
Starry Night and Tree Roots by Van Gogh are in the public domain
Bokeh image is in the public domain
Stylized images copyright Justin Johnson, 2017;
reproduced with permission

Mordvinsev et al, 2015

Gatys et al, 2016

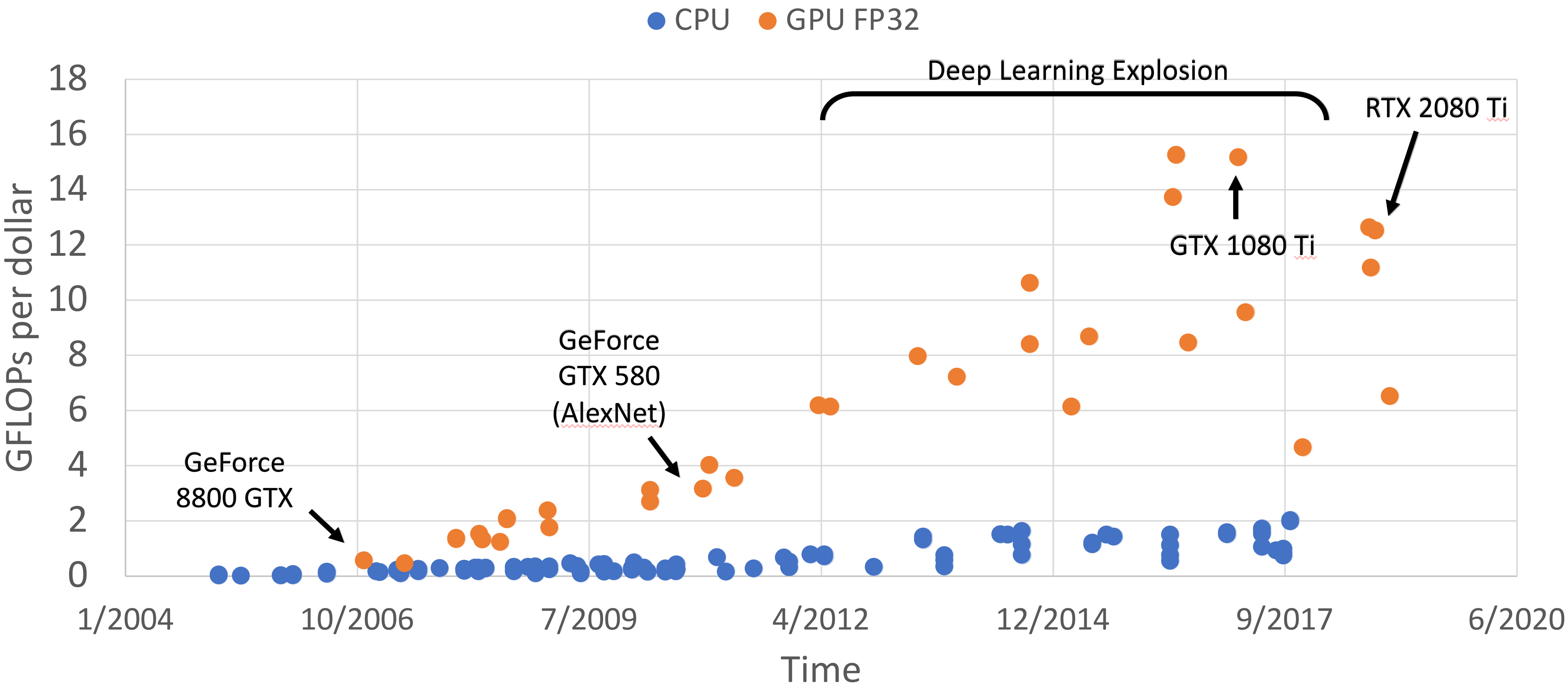


Algorithms

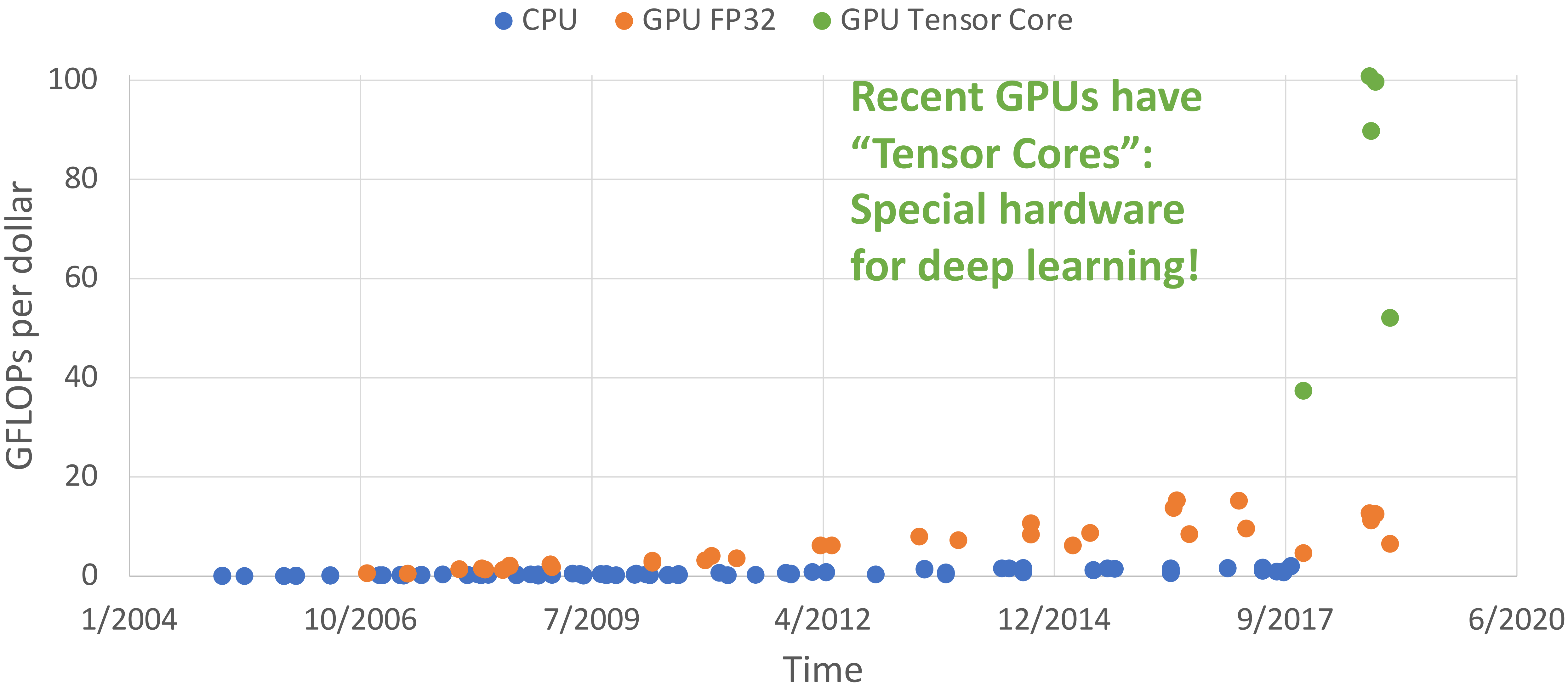
Data

Computation

GFLOPs per Dollar

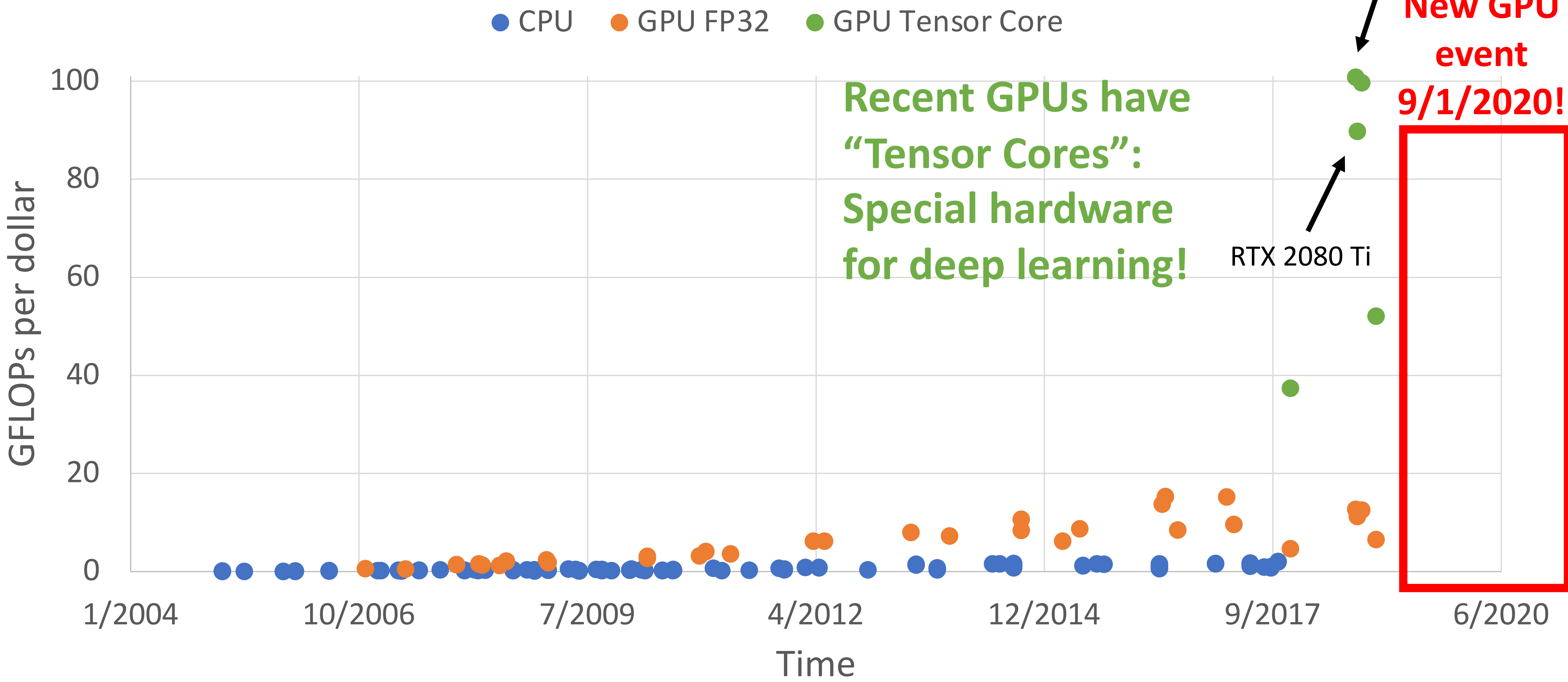


GFLOPs per Dollar



Recent GPUs have "Tensor Cores": Special hardware for deep learning!

GFLOPs per Dollar



Recent GPUs have "Tensor Cores": Special hardware for deep learning!

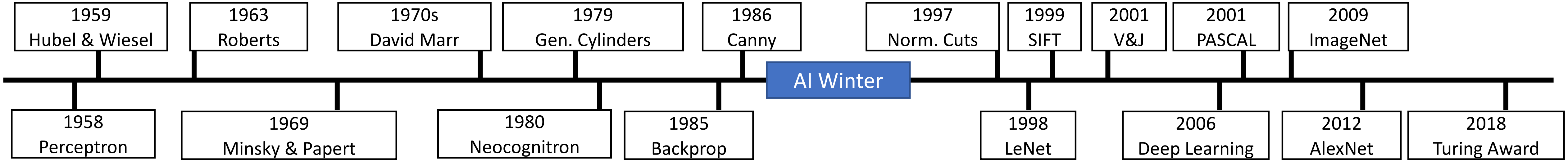
New GPU event 9/1/2020!

2018 Turing Award

Yoshua Bengio

Geoffrey Hinton

Yann LeCun



1958 Perceptron

1959 Hubel & Wiesel

1963 Roberts

1969 Minsky & Papert

1970s David Marr

1979 Gen. Cylinders

1980 Neocognitron

1985 Backprop

1986 Canny

AI Winter

1997 Norm. Cuts

1998 LeNet

1999 SIFT

2001 V&J

2001 PASCAL

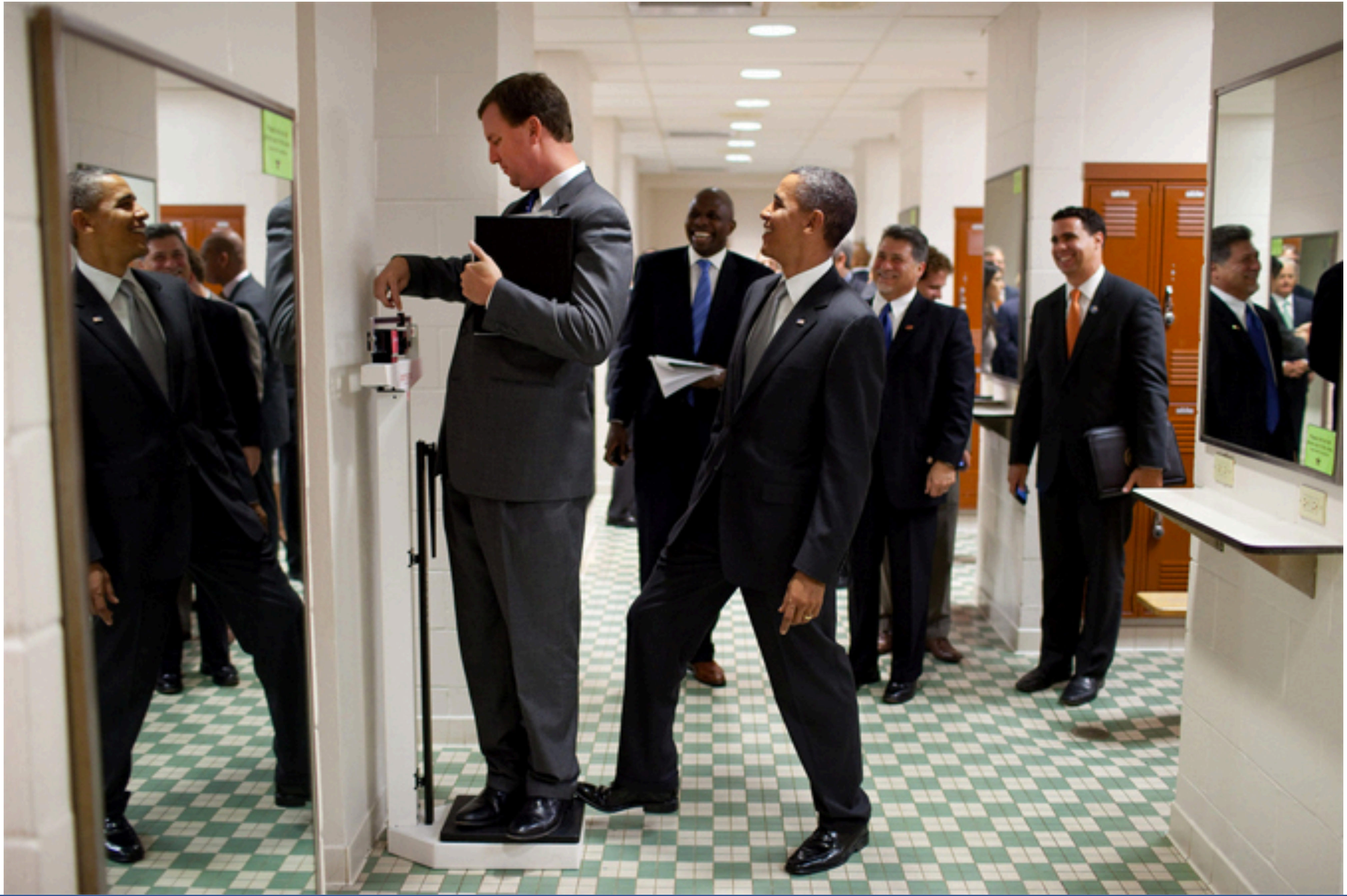
2006 Deep Learning

2009 ImageNet

2012 AlexNet

2018 Turing Award

Despite our success, computer vision still has a long way to go…



Example credit: Andrej Karpathy

Computer Vision Technology Can Better Our Lives

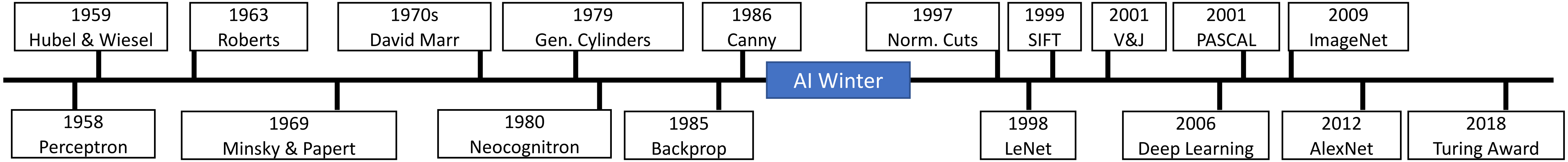
Outside border images, clockwise, starting from top left:
Image by Pop Culture Geek is licensed under CC BY 2.0; changes made
Image by the US Government is in the public domain
Image by the US Government is in the public domain
Image by Glogger is licensed under CC BY-SA 3.0; changes made
Image by Sylenius is licensed under CC BY 3.0; changes made
Image by US Government is in the public domain

Inside four images, clockwise, starting from top left:
Image is CC0 1.0 public domain
Image by Tucania is licensed under CC BY-SA 3.0; changes made
Image by Intuitive Surgical, Inc. is licensed under CC BY-SA 3.0; changes made
Image by Oyundari Zorigtbaatar is licensed under CC BY-SA 4.0

Today’s Agenda

- A brief history of computer vision and deep learning

- Course overview and logistics



1958 Perceptron

1959 Hubel & Wiesel

1963 Roberts

1969 Minsky & Papert

1970s David Marr

1979 Gen. Cylinders

1980 Neocognitron

1985 Backprop

1986 Canny

AI Winter

1997 Norm. Cuts

1998 LeNet

1999 SIFT

2001 V&J

2001 PASCAL

2006 Deep Learning

2009 ImageNet

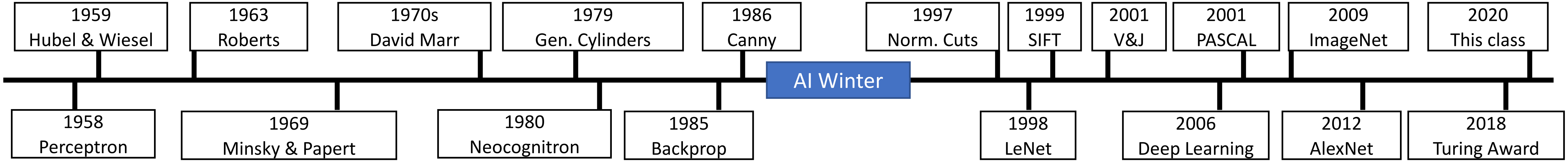
2012 AlexNet

2018 Turing Award

Today's Agenda

- A brief history of computer vision and deep learning

- Course overview and logistics



1958 Perceptron

1959 Hubel & Wiesel

1963 Roberts

1969 Minsky & Papert

1970s David Marr

1979 Gen. Cylinders

1980 Neocognitron

1985 Backprop

1986 Canny

AI Winter

1997 Norm. Cuts

1998 LeNet

1999 SIFT

2001 V&J

2001 PASCAL

2006 Deep Learning

2009 ImageNet

2012 AlexNet

2018 Turing Award

2020 This class

Course Staff

Instructor

Justin Johnson
Assistant Professor, CSE

Graduate Student Instructors

Yunseok Jang

Mohamed El Banani

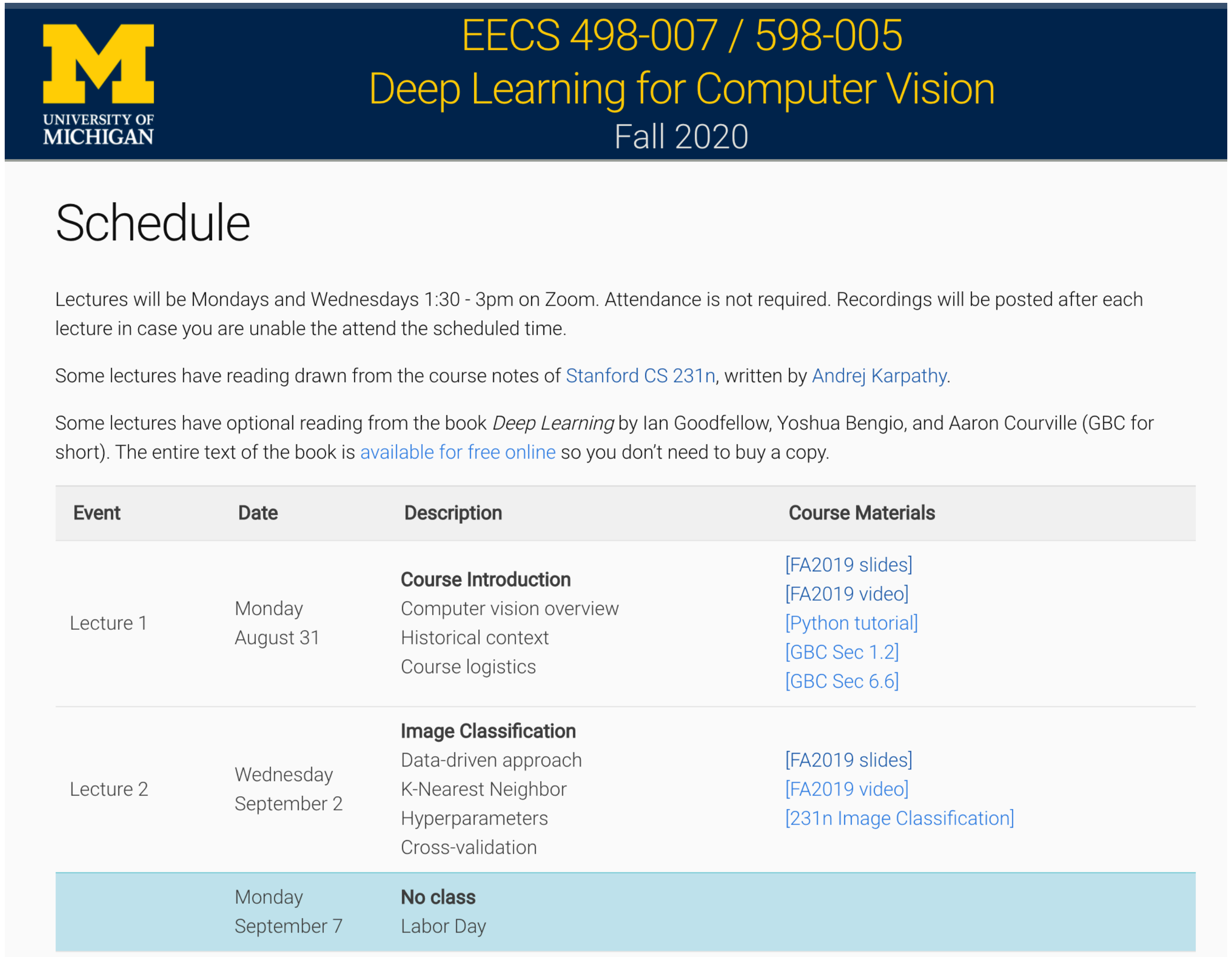
Danish Syed

Yashmeet Gambhir

How to contact us

- Course Website: https://web.eecs.umich.edu/~justincj/teaching/eecs498/
  - Syllabus, schedule, assignments, slides, lecture videos, etc
- Piazza: https://piazza.com/class/ke3a8m6u5wx647
  - (Almost) all questions about the course should go here!
  - We will also use Piazza to communicate with you
  - Use private questions if you want to post code
- EECS Autograder:
  - For turning in homework assignments
  - Still working out details, will update soon
- Google Calendar: For office hours (starting next week)
- Email: Only for sensitive, confidential issues

Course Website: Check the Schedule!



EECS 498-007 / 598-005
Deep Learning for Computer Vision
Fall 2020

Schedule

Lectures will be Mondays and Wednesdays 1:30 - 3pm on Zoom. Attendance is not required. Recordings will be posted after each lecture in case you are unable the attend the scheduled time.

Some lectures have reading drawn from the course notes of Stanford CS 231n, written by Andrej Karpathy.

Some lectures have optional reading from the book *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (GBC for short). The entire text of the book is available for free online so you don't need to buy a copy.

Event	Date	Description	Course Materials
Lecture 1	Monday August 31	**Course Introduction** Computer vision overview Historical context Course logistics	[FA2019 slides] [FA2019 video] [Python tutorial] [GBC Sec 1.2] [GBC Sec 6.6]
Lecture 2	Wednesday September 2	**Image Classification** Data-driven approach K-Nearest Neighbor Hyperparameters Cross-validation	[FA2019 slides] [FA2019 video] [231n Image Classification]
	Monday September 7	**No class** Labor Day	

https://web.eecs.umich.edu/~justincj/teaching/eecs498/FA2020/schedule.html

Piazza Etiquette

- Post only short snippets of code (< 20-30 lines)
- Ask a specific, concrete question
- Explain what you have tried so far, and what happened
- See StackOverflow guide on asking good questions: https://stackoverflow.com/help/how-to-ask

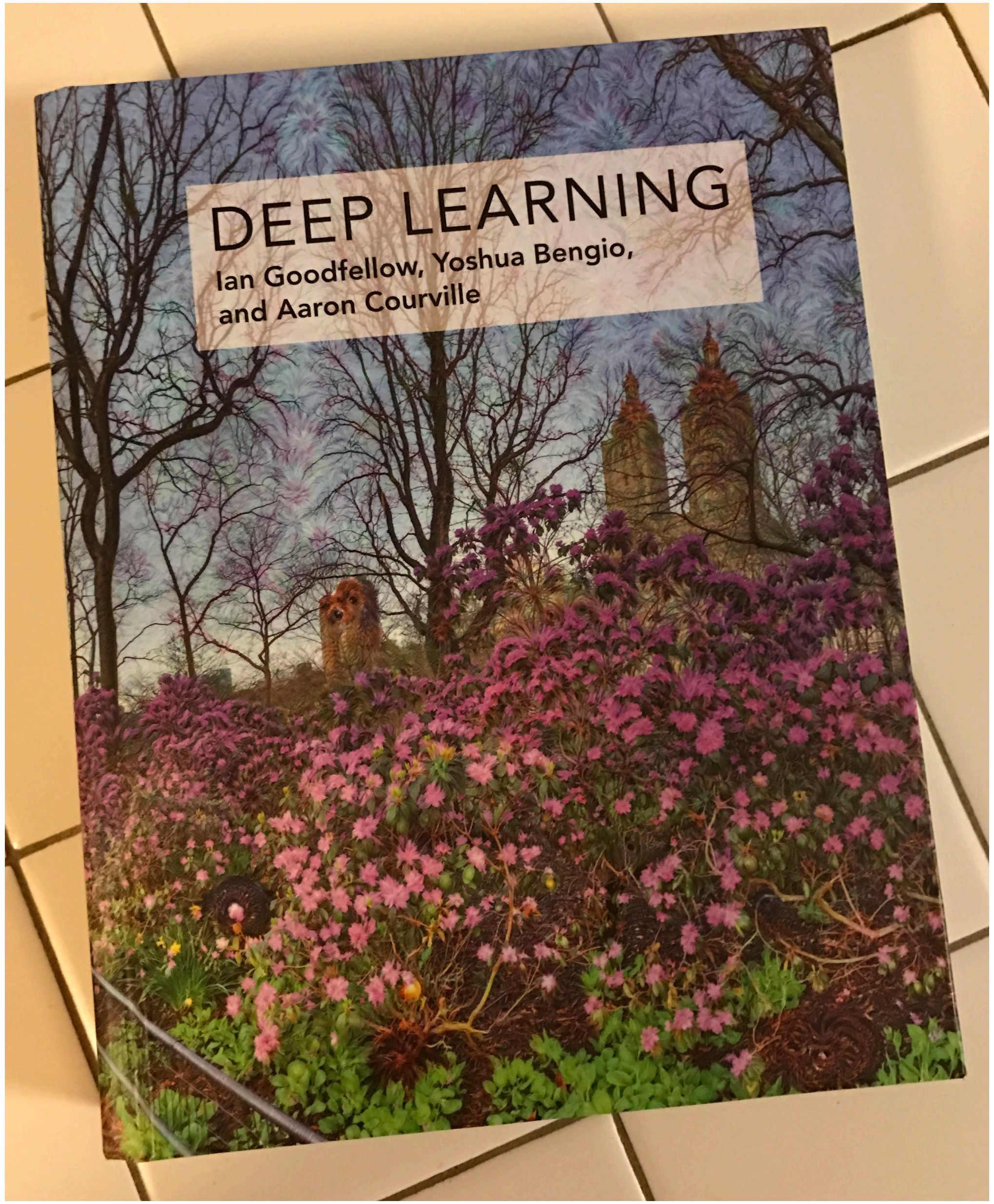
Piazza Etiquette

- Post only short snippets of code (< 20-30 lines)
- Ask a specific, concrete question
- Explain what you have tried so far, and what happened
- See StackOverflow guide on asking good questions: https://stackoverflow.com/help/how-to-ask

- Don’t expect and answer within 30 minutes of posting
- Monday – Friday, 10am – 6pm EST we’ll try to answer within 2 hours
- Other times, we’ll try to answer within 12 hours

Optional Textbook

- *Deep Learning* by Goodfellow, Bengio, and Courville
- Free online



Course Content and Grading

- 6 programming assignments (A1 10%, A2-A6 12%)
  - Homework assignments will use Python, PyTorch, and Google Colab
- Midterm Exam (30%)
- Late policy
  - 3 free late days to use on assignments
  - Once free late days are exhausted, 25% penalty per day

Collaboration Policy

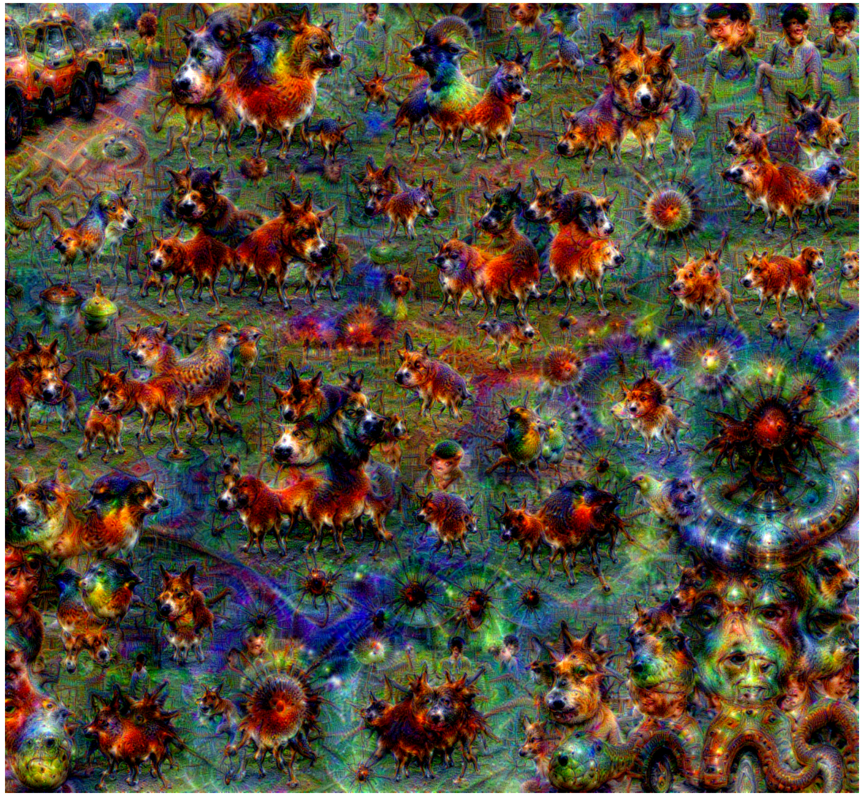
- **Rule 1**: Don’t look at solutions or code that are not your own; everything you submit should be your own work
- **Rule 2**: Don’t share your solution code with others; however discussing ideas or general strategies is fine and encouraged
- **Rule 3**: Indicate in your submissions anyone you worked with
- Turning in something late / incomplete is better than violating the honor code

Course Philosophy

- Thorough and Detailed.
  - This not “Learn PyTorch in 90 days”, nor “Deep Learning in 10 lines of code”
  - Understand how to write from scratch, debug, and train convolutional and other types of deep neural networks
  - We prefer to write from scratch, rather than rely on existing implementations
- Practical
  - Focus on practical techniques for training and debugging neural networks
  - Will use state-of-the-art software tools like PyTorch and TensorFlow
- State of the art
  - Most material we cover is research published in the last 5 years

Course Philosophy

- Will also cover some fun topics:
  - Image captioning
  - DeepDream, Artistic Style Transfer



Course Structure

- First half: Fundamentals
  - Details of how to implement and train different types of networks
  - Fully-connected networks, convolutional networks, recurrent networks
  - How to train and debug, very detailed
- Second half: Applications and “Researchy” topics
  - Object detection, image segmentation, 3D vision, videos
  - Attention, Transformers
  - Vision and Language
  - Generative models: GANs, VAEs, etc
  - Guest Lectures from subject-matter experts
  - Less detailed: provide overview and references, but skip some details

First homework assignment

- Will be released by tomorrow
- Due Friday 9/11/2020
- Next lecture will be enough to complete it

Next time: Image Classification